Calculus, Algebra, and Analysis for JMC

Lectured by Marie-Amelie Lawn, Frank Berkshire

Typed by Aris Zhu Yi Qing

May 13, 2020

Contents

Chapter 1

Group theory

Study of the simplest algebraic structure on a set.

1.1 Basic Definitions and Examples

1.1.1 Binary operations and groups

Definition 1. *Set* is a collection of distinct elements. Let G be a set. ***Binary operation on G*** is a function

$$* : G \times G \to G(\text{Closure is included})$$

Example 2.

- $(\mathbb{N}, +), (\mathbb{Z}, +), (\mathbb{R}, \cdot)$
- $(\mathbb{N}, -)$ not a binary op. Not closed.
- $g, h \in G, g * h = h$
- Find a certain $c \in G$, define $g * h = c \; \forall g, h \in G$

Example 3. Cayley table: Draw a table of all the possible binary operations on a set. How many possible binary operations on a finite set with n elements? In general, there are ∞-many biniary operatiions. In this case, there are n^{n^2} possible binary operations. *In general,* $g_i * g_j \neq g_j * g_i$ (Not commutative!)

Definition 4. A binary operation $*$ on a set G is called associative if

$$(g * h) * k = g * (h * k) \ \forall g, h, k \in G$$

Example 5.

- $+$ on $\mathbb{N}, \mathbb{Z}, \mathbb{R}$? Yes
- $-$ on $\mathbb{R}$? No
- $g * h = g^h$ on $\mathbb{N}$? No

Definition 6. A binary operation is called commutative if

$$\forall g, h \in G, g * h = h * g$$

Example 7.

- $+, \cdot$ on $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$
- matrix multiplication ($AB \neq BA$ in general for A, B in $M(\mathbb{R}^n)$)
- let $g, h \in \mathbb{R}$, $g * h = 1 + g \cdot h$: commutative but *not associative*!

Definition 8. Let $(G, *)$ be a set. An element e is called *left identity* (respectively *right identity*) if:

$$e * g = g (\text{resp. } g * e = g) \ \forall \ g \in G$$

Caution: There might be *many* left/right identities or none.

Example 9.

1. let $(G, *)$ be a set with $g * h := g$. Find the left/right identities.

   ∞-many (or equal to the number of elements) right identities since h satisfies definition $\forall h$. No left identities: wanted $e * g = g = e$ by definitioin of $*$ (*unless only one element*).

2. $(G, *)$, $g * h = 1 + gh$. Ex: No right/left identities.

Idea: We want a good unique identity.

Theorem 10. let $(G, *)$ be set, such that $*$ has both a left identity e_1 *and* a right identity e_2, then

$$e_1 = e_2 =: e \quad \text{and} \quad e \text{ is unique.}$$

Proof.

- $e_1 = e_2$

$$\Rightarrow \left\{ \begin{array}{l} e_1 * g = g \Rightarrow e_1 * e_2 = e_2 \\ g * e_2 = g \Rightarrow e_1 * e_2 = e_1 \end{array} \right\} \forall\, g \in G \Rightarrow \; e_1 = e_2$$

- Unicity: Assume there exists another identity e'.

$$\Rightarrow e' * g = g * e' = g$$

$$e' * g = e' * e = e$$

$$g * e' = e * e' = e'$$

Therefore

$$e = e'.$$

$\square$

As soon as you get one left and one right identity, you have a unique identity e.

Definition 11. let $(G, *)$ be a set. Let $g \in G$. An element $h \in G$ is called left (resp. right) inverse if

$$h * g = e \text{ (resp. } g * h = e).$$

Caution: Again inverses might not exist, there might be many, or *not* the same on both sides.

Example 12.

(1) $(\mathbb{N}, \cdot)$ 1 has an inverse, otherwise *no* inverse.

(2) Find a binary operation on a set of 4 elements with left/right inverses not the same but identity e.

Theorem 13. Let $(G, *)$ be a set with associative binary operation and identity e. Then if h_1 is left inverse, and h_2 is right inverse, then

$$h_1 = h_2 = g^{-1} \text{ and } \textit{it is unique}.$$

Proof.

- $h_1 = h_2$

  $h_1 * g = e, g * h_2 = e$. Therefore

  $$h_2 = e * h_2 = (h_1 * g) * h_2 = h_1 * (g * h_2) = h_1 * e = h_1$$

- unicity: Assume $\exists g'^{-1}$ another inverse.

  $$g'^{-1} = e * g'^{-1} = (g^{-1} * g) * g'^{-1} = g^{-1} * (g * g'^{-1}) = g^{-1} * e = g^{-1}$$

$\square$

(Group) Definition 14. A set $(G, *)$ with binary operatioin $*$ is called a *group* if:

(1) $*$ is associative

(2) $\exists e \in G$ an identity $\forall g \in G$

(3) All elements $g \in G$ have an inverse g^{-1}

<u>Attention</u>: The identity and inverses are *unique* by our previous results.

Example 15.

- $(\mathbb{Z}, +), (\mathbb{Z}_n, +)$ (will see this later) are groups.
- $(\mathbb{N}, +)$ not a group $\Rightarrow$ no inverses.
- $(\mathbb{C}, \cdot)$ not a group (0 has no multiplicative inverse), but $(\mathbb{C}^*, \cdot)$ is. ($\mathbb{C}^* = \mathbb{C}\backslash\{0\}$)
- $(G = \{e\}, *)$ with $e * e = e$ is a group called the *trivial group*.
- Empty set $\varnothing$ is not a group (No identity element.)

Definition 16. Let G be a group. It is called finite if it has finitely many elements.

Notation: $|G| = n$ (number of elements)

We say that G has ***order*** n. If $|G| = \infty$, the G is called an infinite group.

Example 17.

- the trivial group is finite, $|G| = 1$
- let $G = \{1, -1, i, -i\} \subset \mathbb{C}$, with $* = \cdot$. Is it a group? Yes. Check associativity, identity, and inverses.

(Abelian Group) Definition 18. A group is called *Abelian* if $*$ is commutative.

Example 19.

- previous example, trvial group, $(\mathbb{Z}, +), (\mathbb{C}^*, \cdot)$
- let $GL(\mathbb{R}^n)$ be the set of all invertible $n \times n$ matrices, $*$ = matrix multiplication. It is associative: $(AB)C = A(BC)$; It has identity: I_n. It has inverses: yes since we asked for it. So this is a group of matrices. But this is not Abelian since $AB \neq BA$.
- let G be the set of *invertible* functions with $* = \circ$, the composition of functions. Identity is $F(x) = x$; they are associative, invertible, but *not Abelian*.

1.1.2 Consequences of the axioms of group

Theorem 20. Let $(G, *)$ be a group, $g, h \in G$. Then

$$(g * h)^{-1} = h^{-1} * g^{-1}$$

Proof. To show: $(g * h) * (h^{-1} * g^{-1}) = e$.

Using assocativity, we have

$$g * (h * h^{-1}) * g^{-1} = g * g^{-1} = e$$

□

Definition 21. Let $n \in \mathbb{Z}$, let $(G, *)$ be a group and let $g \in G$. Then we definie g^n as follows:

$$g^n = \begin{cases} g * g * \cdots * g & n > 0 \\ g^{-1} * g^{-1} * \cdots * g^{-1} & n < 0 \\ e & n = 0 \end{cases}$$

where in the first case there are n copies of g in the product and ni the second there are $-n$ copies of g^{-1}, so that $g^n = (g^{-1})^{-n}$.

Theorem 22. Let $n, m \in \mathbb{Z}$ and let $G, *$ be a group. Then

1. $g^n * g^m = g^{n+m}$
2. $(g^n)^m = g^{nm}$

Proof. Exercise! (Hint: Induction.) □

1.1.3 Modular Artihmetic and the group $\mathbb{Z}_n$

Definition 23. let $n > 0$, $n \in \mathbb{Z}$ fixed, $a, b \in \mathbb{Z}$. a and b are called ***congruent modulo*** n if $n|a - b$.

Definition 24. $\forall a, b, c \in \mathbb{Z}$, $n > 0$ fixed in $\mathbb{Z}$:

(1) $a \equiv a \mod n$ (reflexivity)

(2) If $a \equiv b \mod n \iff b \equiv a \mod n$ (symmetry)

(3) if $a \equiv b \mod n$ and $b \equiv c \mod n \implies a \equiv c \mod n$ (transitivity)

Definition 25. Given a set S and an equivalence relation $\sim$ on S, the ***equivalence class*** of an element a in S is the set $\{x \in S \mid x \sim a\}$.

Definition 26. Define the equivalence class of $a \in \mathbb{Z}$ in the relation of congruence modulo n as:

$$[a]_n := \{b \in \mathbb{Z} \,|\, b \equiv a \mod n\}$$

Definition 27. Define equivalence classes $\mathbb{Z}_n$ as

$$\mathbb{Z}_n := \{[0]_n, [1]_n, \ldots [n-1]_n\}$$

with 2 binary operations on $\mathbb{Z}_n$:

$+$: $\mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$, $([a]_n, [b]_n) \mapsto [a+b]_n$

$\cdot$: $\mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$, $([a]_n, [b]_n) \mapsto [ab]_n$

As we can see from the following lemma, the two operations are well-defined.

Lemma 28. Let $a, a', b, b' \in \mathbb{Z}$ s.t. $[a]_n = [a']_n, [b]_n = [b']_n$. Then $[a+b]_n = [a'+b']_n$, $[a \cdot b]_n = [a' \cdot b']_n$.

Proof. Exercise! □

Theorem 29. $(\mathbb{Z}_n, +)$ is an Abelian group.

Proof.

(1) Associativity:

$$\begin{aligned}
([a]_n + [b]_n) + [c]_n &= [a+b]_n + [c]_n \\
&= [a+b+c]_n \\
&= [a]_n + [b+c]_n \\
&= [a]_n + ([b]_n + [c]_n)
\end{aligned}$$

(2) Commutativity:

$$\begin{aligned}[a]_n + [b]_n &= [a+b]_n \\ &= [b+a]_n \\ &= [b]_n + [a]_n\end{aligned}$$

(3) Identity element: $[0]_n$

(4) Inverse: Any element $[a]_n$ has an inverse $[-a]_n$.

$\square$

Example 30. $(\mathbb{Z}_n, \cdot)$ is an Abelian group?
Similary to above for associative, commutative, and identity.
<u>Inverses</u>:

Draw Caley table for $(\mathbb{Z}_3, \cdot)$. We realize that $[0]_3$ has no inverses. But $(\mathbb{Z}_3 \backslash \{[0]_3\}, \cdot)$ is.

Similarly, for $(\mathbb{Z}_4, \cdot)$, it does not have inverses for all classes.

<u>Caution</u>: In general $(\mathbb{Z}_n, \cdot)$ is *not* a group. The idea then is to make it a group by removing non-invertible elements.

Lemma 31. The element $[a]_n \in \mathbb{Z}_n$ has an inverse $\iff$ $(a, n) = 1$.

Proof. $(a, n) = 1 \iff \exists b, c \in \mathbb{Z}$, s.t $ab + cn = 1 \iff cn = 1 - ab \iff \exists [b]_n$ s.t. $[a]_n[b]_n = [1]_n$. $\square$

Definition 32. $\mathbb{Z}_n^* := \{[a]_n \in \mathbb{Z}_n \mid \exists b \in \mathbb{Z} \text{ s.t. } [a]_n[b]_n = [1]_n\}$.

Theorem 33. $(\mathbb{Z}_n^*, \cdot)$ is an Abelian group.

Proof. To Show: if $[a]_n, [b]_n \in (\mathbb{Z}_n^*, \cdot) \Rightarrow [a]_n \cdot [b]_n \in (\mathbb{Z}_n^*, \cdot)$.
$\Rightarrow$ $(a, n) = (b, n) = 1 \Rightarrow$ $(ab, n) = 1 \Rightarrow$ $[ab]_n$ has inverse $[a]_n[b]_n$.

Alternatively: if g, h have inverse, $h^{-1}g^{-1}$ is inverse of gh. $\square$

1.2 Cyclic groups

Definition 34. Let G be a group, $g \in G$. The ***order*** of g is the *smallest positive* integer $n > 0$ such that $g^n = e$.

Notation: ord $g = n$. If $n = \infty$, then g is called of infinite order.

Example 35. $G = (\mathbb{C}^*, \cdot)$, ord $(-1) = 2$, ord $i = 4$, ord $2 = \infty$

Lemma 36. Let G be a finite group. Then every element $g \in G$ has finite orders.

Proof. Assume $g \in G$ has infinite orders. Write the list: $g^0, g^1, g^2, \ldots$

Since $|G| = n < \infty$, there are two elements g^k, g^l s.t. $g^k = g^l$, $k > l$. $\iff$ $g^k g^{-l} = e \iff g^{k-l} = e$.

But then ord $g \leq k - l < \infty$. □

Lemma 37. Let G be a group, $g \in G$, ord $g = n$. Then all elements $\{g^0, g^1, g^2, \ldots, g^{n-1}\}$ are distinct.

Proof. Assume that $g^i = g^j$ for some $i, j, 0 \leq i \leq j \leq n-1$. Then $g^{j-i} = g^0 = e$ and $j - i < n$. Since n is the smallest integer s.t. $g^n = e$, contradicts with the condition. □

Corollary 38. If $|G| = n < \infty$, $g \in G$, then ord $g \leq n$.

Proof. Let $i \in \mathbb{Z}, i \geq n+1$, and assume $\exists g^i = e$, where $g \in G$, i is the smallest such integer. By previous lemma, $\{g_0, g_1, g_2, \ldots, g^{i-1}\}$ all distinct. There are i elements $i > n = |G|$, contradiction. □

Definition 39. We call a group G ***cyclic*** if

$$\exists g \in G \text{ s.t. } G = \{g^n | n \in \mathbb{Z}\}.$$

g is called a ***generator***.

Example 40.

- $(\mathbb{Z}, +)$. $2 = 1^2 = 1 + 1$, $n = 1^n$.
- $(\mathbb{Z}_n, +)$, generator $[1]_n$.
- $\{\pm 1, \pm i\}$, generator $\pm i$.

Lemma 41. All cyclic groups are Abelian.

Proof. To show: $\forall h, k \in G, h \cdot k = k \cdot h$.
G is cyclic $\Rightarrow G = \{g^n | n \in \mathbb{Z}\}$ for some generators $g \in G \Rightarrow h = g^i, k = g^j$.
$\Rightarrow h \cdot k = g^i \cdot g^j = g^{i+j} = g^{j+i} = g^j \cdot g^i = k \cdot h$. □

Warning: The converse *is not* true (Abelian does not imply cyclic) One counter example is $(\mathbb{Q}, +)$. Assume $\mathbb{Q}$ is cyclic under $+$.

$$\Rightarrow \exists g \in \mathbb{Q} \text{ s.t. } q = g^n (= ng) \forall q \in \mathbb{Q}.$$

Take $\frac{g}{2}$ ($\in \mathbb{Q}$ since $g \in \mathbb{Q}$)

$$\Rightarrow \frac{g}{2} = ng \text{ for some } n \in \mathbb{Z}.$$

contradicting with original statements.

Lemma 42. Let G be a *finite* group, $|G| = n$. So

$$G \text{ is cyclic } \iff G \text{ contains an element of order } n$$

Proof.
"$\Rightarrow$": G is cyclic $\Rightarrow G$ has generator g. Assume ord $g = k$, so

$$\left\{g^0, \ldots, g^{k-1}\right\} \text{ are distinct.}$$

$\Rightarrow k = n$ since $|G| = n$.
"$\Leftarrow$": Let assume $\exists g \in G$, ord $g = n$.

$$\Rightarrow \left\{g^0, g^1, \ldots, g^{n-1}\right\} \text{ are all distinct.}$$

But $|G| = n$, hence g generates all the group. □

Lemma 43. Let G be a finite group. Then if G is cyclic, it has at most one element of order 2.

Proof. Since G is finite ($|G| = n$), and cyclic, $\exists g \in G$ of order n ($g^n = e$), and $G = \{g^0, g^1, \ldots, g^{n-1}\}$. Assume $\exists$ an element of order 2: $h = g^i, (i \geq 0, i \in \mathbb{Z})$, then

$$\left(g^i\right)^2 = e = g^{2i} \Rightarrow 2i = n \Rightarrow \begin{cases} n \text{ is even: exactly one element,} \\ n \text{ is odd: no element of order 2.} \end{cases}$$

Subtlety: $2i$ cannot be equal to $kn, k \in \mathbb{N}\backslash\{0,1\}$, whereas $3i$ or mi, $m \geq 3$, is possible to be equated to kn, thus no conclusion can be drawn for elements of order greater than 2! □

Example 44. Are $(\mathbb{Z}_5^*, \cdot), (\mathbb{Z}_{15}^*, \cdot)$ cyclic? (Recall that the notation $\mathbb{Z}^* = \mathbb{Z}\backslash\{0\}$, and $\mathbb{Z}_n^*$ = set of all invertible congruence classes $[a]_n$.) (Hint: Use the previous lemma, or find out the generator.)

1.3 Symmetric groups

1.3.1 Permutations

Definition 45. A function f from a set X to a set Y is called

- ***one-to-one*** or ***injective*** if $f(x_1) = f(x_2) \Rightarrow x_1 = x_2 \, \forall x_1, x_2 \in X$.
- ***onto*** or ***surjective*** if $\forall y \in Y, \exists x \in X$ s.t. $f(x) = y$.
- a ***bijection*** if it is both *injective* and *surjective*.

Furthermore, f is a bijection iff there is an inverse function $g : Y \mapsto X$ s.t. $g \circ f$ is the identity function on X and $f \circ g$ is the identity function on Y.

Definition 46. A ***permutation*** is a bijective function:

$$\sigma : \{1, 2, \ldots, n\} \mapsto \{1, 2, \ldots, n\}.$$

Notation: We write the permutation as *two-row notation*: we write down the numbers 1 to n, and underneath each number i we write down the number that σ sends i to:

$$\begin{vmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{vmatrix}$$

Because σ is a bijection, the bottom row of the table consists of the numbers $1, 2, \ldots, n$ in some order. So a permutation is a 're-ordering' of the numbers 1 to n.

Definition 47. The set of all permutation $S_n :=$ $\{\sigma : \{1, 2, \ldots, n\} \mapsto \{1, 2, \ldots, n\}\}$ is called the ***symmetric group*** (on n symbols).

Theorem 48. The set $(S_n, \circ)$ is a group.

Proof.

- Closure: Let $\nu, \tau \in S_n$, then ν, τ are bijective by definition, so are $\tau \circ \nu$ and $\nu \circ \tau$.
- Associativity: composition of functions is associative.
- Identity: identity $\nu(h) = k \ \forall k \in \{1, 2, \ldots, n\}$.
- Inverses: By definition: bijections $\iff$ $\exists$ inverses!

□

Theorem 49. $(S_n, \circ)$ is not Abelian.

Proof. Exercise! □

Proposition 50. $|S_n| = n!$.

Proof. Exercise! □

1.3.2 Cycle

Definition 51. A permutation is called a ***cycle*** if there is **a** sequence $\{a_1, a_2, \ldots, a_k\}$ of distinct numbers s.t.

$$\sigma(a_1) = a_2, \quad \sigma(a_2) = a_3, \quad \ldots, \quad \sigma(a_{k-1}) = a_k, \quad \sigma(a_k) = a_1$$

and $\sigma(i) = i$ for any other i *not* in the sequence. The number k is called the ***length*** of the cycle, and we often abbreviate 'cycle of length k' to 'k-***cycle***'.

Example 52.

$$\nu = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{vmatrix} \quad \text{and} \quad \tau = \begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{vmatrix}$$

ν is a 3-cycle, it rotates the numbers 1, 2, 3 and fixes 4. τ is not a cycle: no numbers are fixed, so if it was a cycle it would have to be 4-cycle, but it is not.

Proposition 53. The order of a k-cycle is k.

Proof. We know immediately that $\sigma^k = \text{id}$ by definition. $\Rightarrow \text{ord}\,\sigma \leq k$.

Assume that $\text{ord}\,\sigma = i < k$. But by definition of $\sigma^i(a_1) = a_{i+1} \neq a_1$. □

Notation of a k-cycle: $(a_1, a_2, \ldots, a_k)$. This means sending $a_1 \mapsto a_2 \mapsto a_3 \mapsto \cdots \mapsto a_k \mapsto a_1$ and fixes all other elements. This only makes sense if the numbers $a_1, a_2, \ldots, a_k$ are all distinct (or this permutation would not be a cycle).

Example 54. From the previous example, we would write the 3-cycle ν as (1, 2, 3).

Note:

(1) There are several different ways of writing the same cycle, for instance (1, 2, 3), (2, 3, 1), (3, 1, 2) are all the same. The usual convention is to put the smallest number first.

(2) A cycle of length one has to be the identity permutation. So the 1-cycles (1), (3), (42), all denote the identity. The usual convention is to use (1), and this makes sense in any S_n.

(3) Cycles make sense if all elements are distinct.

Example 55. The permutation $\tau \in S_4$ from the second previous example is not a cycle, but it is easy to see that it can be expressed as the composition

$$\tau = (3,4)(1,2)$$

of two 2-cycles.

Definition 56. Two cycles $(a_1, a_2, \dots, a_k), (b_1, b_2, \dots, b_m)$ are ***disjoint*** if no a_i is equal to any b_j.

Theorem 57. Disjoint cycles commute if the two cycles are disjonit, i.e. if α, β are disjoint cycles of the set $\{1, 2, \dots, n\}$, then $\alpha \circ \beta = \beta \circ \alpha$.

Proof. Exercise! □

Lemma 58. Let $\sigma \in S^n$ be a permutation.

1. For any $i \in \{1, \dots, n\}$, there is a positive integer d such that $\sigma^d(i) = i$. (In fact, such smallest $d \in [1, n]$.)
2. If d is the smallest positive integer such that $\sigma^d(i) = i$, then the numbers $i, \sigma(i), \sigma^2(i), \dots, \sigma^{d-1}(i)$ are all distinct.
3. If $j \in \{1, \dots, n\}$ is not in the set $\{i, \sigma(i), \dots, \sigma^{d-1}(i)\}$, then neither is $\sigma(j)$.

Proof. Exercise! □

Proposition 59. Any permutation can be expressed as a product of some number of disjoint cycles.

Proof. The proof is given by an explicit algorithm. Pick any $\sigma \in S_n$. Then pick any number $i \in \{1, \ldots, n\}$. By the previous lemma, there is an integer d such that $\sigma^d(i) = i$. Take the smallest such d, and also by previous lemma that $i, \sigma(i), \ldots, \sigma^{d-1}(i)$ are all distinct, we can then form the cycle

$$(i, \sigma(i), \ldots, \sigma^{d-1}(i))$$

Repeat the above process by choosing an element which does not occur in the cycle until all numbers are in one of the cycles. The permutation σ will be the product of our list of cycles. $\square$

Definition 60. When σ is factored into disjoint cycles $\gamma_1\gamma_2 \ldots \gamma_r$ we can record the lengths $(k_1, k_2, \ldots, k_r)$ of the cycles that occur, and the list is called the ***cycle-type*** of σ.

Example 61. Factor and find the cycle-type of

$$\sigma = \begin{vmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 1 & 3 & 2 & 6 & 7 & 5 \end{vmatrix}.$$

Answer: $\sigma = (1, 4, 2)(5, 6, 7)$, and the cycle-type of σ is $(3, 3)$. (We can leave out the 1's from the list, they are not important.)

1.4 subgroup

Definition 62. Let $(G, *)$ be a group. $H \subseteq G$ a subset. Then H is called a subgroup of G if:

1. $\forall g, h \in H, g * h \in H$. (Closure)
2. $e \in G$ is also in H. (identity element)
3. $g \in H \Rightarrow g^{-1} \in H$. (inverses)

Note: We can replace (2) with (2') $H \neq \varnothing$.

Proof. $H \neq \varnothing \iff \exists h \in H \Rightarrow h^{-1} \in H \Rightarrow h * h^{-1} = e \in H$. □

Notation: $H \leq G$ means H is a subgroup of G. v.s. $\subseteq$.

Example 63.
- $(\mathbb{Z}, +) \leq (\mathbb{Q}, +) \leq (\mathbb{R}, +) \leq (\mathbb{C}, +)$.
- $n\mathbb{Z} := (\{nz | z \in \mathbb{Z}\}, +) \leq (\mathbb{Z}, +)$.
- Any group has two immediate subgroup: $(G, *) \leq (G, *)$, and $(\{e\}, *)$ trivial subgroup. If $H \leq G$, $H \neq G$, G is called *proper*; if $H \neq \{e\}$, H is called *non-trivial*.

Proposition 64. Let $(G, *)$ be a group, $H \subseteq G$, $H \neq \varnothing$. Then if $\forall x, y \in H, x * y^{-1} \in H \Rightarrow H \leq G$.

Proof. To show: H is subgroup.

1. $H \neq \varnothing \Rightarrow \exists x \in H$, take $y = x$ (by assumption) $\Rightarrow x * y^{-1} = x * x^{-1} = e \in H$.
2. Inverse: Assume $x \in H$, set $y = x$, and the other as the identity: (by assumption) $\Rightarrow e * x^{-1} = x^{-1} \in H$.
3. Closure: Take $x, y \in H$, we know that by the previous point, $y^{-1} \in H$. By assumption, $x * (y^{-1})^{-1} = x * y \in H$.

□

Example 65. Show that $H = \{\sigma \in S_n | \sigma(1) = 1\} \leq S_n$ using subgroup test.

- $H \neq \varnothing$ since $\mathrm{id}(i) = i \forall i \in \{1, \ldots, n\} \Rightarrow \mathrm{id}(1) = 1$, hence $\mathrm{id} \in H$.
- Take $\sigma, \tau \in H$. To show $\sigma \circ \tau^{-1} \in H \iff \sigma \circ \tau^{-1}(1) = 1 \Rightarrow \sigma(1) = 1$. Therefore $\sigma \circ \tau^{-1} \in H \leq S_n$.

Definition 66. Let $(G, *)$ be a group, $g \in G$, $\langle g \rangle = \{g^i | i \in \mathbb{Z}\}$. Then $\langle g \rangle$ is called the ***cyclic subgroup*** of G generated by g.

Proposition 67. $\langle g \rangle \leq G$.

Proof. Subgroup test:

- To show $\langle g \rangle \neq \varnothing$.
- Pick $x, y \in \langle g \rangle \Rightarrow x = g^i, y = g^j$. Now $x * y^{-1} = g^i g^{-j} \in \langle g \rangle$.

$\square$

Lemma 68. If ord $g = n$, then $|\langle g \rangle| = n$.

Proof. ord $g = n \Rightarrow \{g^0, g^1, g^2, \ldots, g^{n-1}\}$ all distinct. $\Rightarrow |\langle g \rangle| \geq n$. To show $|\langle g \rangle| = n$. Take $i \in \mathbb{Z}, i \geq n$. By the Euclidean algorithm: $i = qn + r$ for some $q, r \in \mathbb{Z}, 0 \leq r < n$. Now any element $g^i = g^{qn+r} = g^{qn} \cdot g^r = eg^r = g^r$. So any element of $\langle g \rangle$ is one of the list $\{g^0, g^1, \ldots, g^{n-1}\} \Rightarrow |\langle g \rangle| = n$. $\square$

Example 69.

$$\sigma = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{vmatrix} \in S_3$$

So ord $\sigma = 3$. $\langle \sigma \rangle = \{e, (1,2,3), (1,3,2)\}$.

1.5 Cosets and Lagrange Theorem

Definition 70. Let $(G, *)$ be a group. $H \leq G$, $g \in G$.

- The ***left coset*** of H by g is $gH := \{gh | h \in H\}$.
- Similarly, the ***right coset*** of H by g is $Hg := \{hg | h \in H\}$.

Notation: Set of left cosets: $G : H := \{gH | g \in G\}$. Set of right cosets: $H : G := \{Hg | g \in G\}$.

Warning: If G is Abelian, $gH = Hg \forall g$.

Example 71. Take again: $\langle (1,2,3) \rangle \leq S_3$. Compute the left and right coset of $(1,2)$ and $(2,3)$.

Proposition 72. Let $(G, *)$ be a group, $H \leq G$, $g_1, g_2 \in G$. Then $g_1H = g_2H \iff g_2 \in g_1H$.

Proof.

- "$\Rightarrow$ "

  Assume $g_1H = g_2H$, $e \in H \Rightarrow g_2e \in g_2H = g_1H$.

- "$\Leftarrow$"

  $g_2 \in g_1H \iff \exists h \in H$ s.t. $g_2 = g_1h$.

  First $g_1H \leq g_2H$. An element of g_1H is of the form g_1h_1 for $h_1 \in H$.

  $$\Rightarrow g_1h_1 = (g_2h^{-1})h_1 = g_2(h^{-1}h_1) \in g_2H.$$

  Now $g_2H \leq g_1H$.

  Any element of g_2H is of the form $g_2h_2 = (g_1h)h_2 = g_1(hh_2) \in g_1H$.

□

Corollary 73. Every element $g \in G$ lies in exactly one of the left cosets of H.

Proof. Exercise! □

Definition 74. The left cosets form a ***partition*** of G, they are a collection of subsets $g_1H, g_2H, \ldots \subset G$ such that

$$G = \bigcup g_iH$$

and the intersection of any two of these subsets is empty.

Example 75. Consider the group $(\mathbb{Z}_6, +)$, and the cyclic subgroup

$$H = \langle [3] \rangle = \{[0], [3]\}.$$

The cosets of H are

$$[0] + H = H = \{[0], [3]\} = [3] + H$$

$$[1] + H = \{[1], [4]\} = [4] + H$$
$$[2] + H = \{[2], [5]\} = [5] + H$$

This group is Abelian so there is no distinction between left and right cosets. Notice that all three cosets have the same size.

Lemma 76. Let G be a group and let $H \leq G$ be finite. Then all left cosets of H have the same size, i.e.

$$|gH| = |H| \; \forall g \in G.$$

Proof. Exercise! (Hint: bijection!) □

(Lagrange) Theorem 77. Let G be a finite group and $H \leq G$, then

$$|G| = |H| \cdot |G : H|,$$

in particular the order of H divides the order of G.

Corollary 78. Let G be a finite group and let $g \in G$. Then $\operatorname{ord} g \mid |G|$.

Proof. Exercise! □

Extension: if $g \in G$ is any element of a finite group, then

$$g^{|G|} = e.$$

Corollary 79. Let G be a finite group of size p, where p is a prime number. Then G is cyclic.

Proof. Exercise! □

(Fermat's Little Theorem) Corollary 80. Let $a \in \mathbb{Z}$ and let p be a prime number. If $a \not\equiv 0 \mod p$ then

$$a^{p-1} \equiv 1 \mod p.$$

Proof. Exercise! (Hint: Use $(\mathbb{Z}_p^*, \times)$ together with Lagrange theorem.) (Reminder: $\mathbb{Z}_p^*$ means invertible set of equivalence classes of p.) □

1.6 Future Direction of Study in Group Theory

1. "Normal Subgroup" $\rightarrow$ "Simple Group"
2. number of subgroups in a group $\rightarrow$ "Sylow theorems" $\rightarrow$ "Galois Theory"
3. study of symmetries $\rightarrow$ "Lie Groups"

Chapter 2

Applied Mathematical Methods

2.1 Differential Equations

2.1.1 Definitions and examples

Definition 81. An ***ordinary differential equation*** (ODE) for $y(x)$ is an equation involving derivatives of y.

$$f(x, y, \frac{\mathrm{d}y}{\mathrm{d}x}, \frac{\mathrm{d}^2y}{\mathrm{d}x^2}, \ldots, \frac{\mathrm{d}^ny}{\mathrm{d}x^n}) = 0 \tag{2.1}$$

$$\frac{\mathrm{d}^ny}{\mathrm{d}x^n} = F(x, y, \frac{\mathrm{d}y}{\mathrm{d}x}, \ldots, \frac{\mathrm{d}^{n-1}y}{\mathrm{d}x^{n-1}})$$

and we seek a solution (or solutions) for $y(x)$ satisfying the equations. (If there are more independent variables then we have a partial differential equation (PDE).)

Definition 82.

Order is the order of the highest derivative present.

Degree is the power of the highest derivative when fractional powers have been removed.

Linear *differential equation* is a differential equation that is defined by a *linear polynomial* in the unknown function and its derivative in each term of equation(2.1).

Example 83.

(a) Particle moving along a line with a given force $\to x(t)$ position as function of time t.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = f\left(t, x, \frac{\mathrm{d}x}{\mathrm{d}t}\right)$$

e.g.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\omega^2 x - 2k\frac{\mathrm{d}x}{\mathrm{d}t}$$

The first term is regarding the restoring force, while the second term is regarding the damping/friction. The function is of order 2, degree 1, and linear.

(b) Radius of curvature of a curve

It can be shown that

$$R(x, y) = \frac{\left[1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{\frac{3}{2}}}{\frac{\mathrm{d}^2y}{\mathrm{d}x^2}}$$

The function is of order 2 and degree 2.

(c) Simple growth and decay

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = kQ$$

The function is of order 1, degree 1, and linear. e.g.

(1) $k > 0$. Q as the quantity of money, and $k = (1 + \frac{r}{100})$, and r being the rate of interest.

(2) $k < 0$. Q as the amount of radioactive material, and k as the decay rate.

Hence, obviously $Q(t) = Q_0e^{kt}$ where $Q_0 = Q(0)$ at $t = 0$.

(d) Population dynamics

$P(t)$ as population over time and $F(t)$ as food over time, with

$$\frac{\mathrm{d}P}{\mathrm{d}t} = aP(a > 0) \tag{2.2}$$

$$\frac{\mathrm{d}F}{\mathrm{d}t} = c(c > 0)$$

These two equations form a linear system, with both being of order 1, degree 1.

So $P(t) = P_0 e^{at}, F(t) = ct + F_0$. Misery! Population outgrows food supply.

Pierre Verhulst (1845) replaced a in equation(2.2) with $(a - bP)$ so that growth decreases as P increases:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = aP - bP^2 \tag{2.3}$$

This is in fact a *logistic ODE*, with order 1, degree 1, and nonlinear.

Note: Equation(2.3) is *separable*. Alternatively we can note that equation(2.3) is an example of a *Bernoulli differential equation*

$$\frac{\mathrm{d}y}{\mathrm{d}x} + F(x)y = H(x)y^n \tag{2.4}$$

with $n \neq 0, 1$ Substitution on $z(x) = (y(x))^{1-n} \Rightarrow$ a *linear* equation for $z(x) \to$ solution. (See below)

(e) Predator-Prey System

$x(t)$ as prey and $y(t)$ as predators, we have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = ax - bxy, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -cy + dxy \tag{2.5}$$

Note: Equation(2.5) is *separable* when written in principle

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}} \Rightarrow y(x) \Rightarrow x(t), y(t)$$

This is of order 1, degree 1, and a nonlinear system.

(f) Combat Model System

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -ay, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -bx \tag{2.6}$$

This is of order 1, degree 1, and linear system.

Note: Again equation(2.6) is *separable* when written as $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{bx}{ay} \Rightarrow y(x) \Rightarrow x(t), y(t)$

In general the solution of a differential equation of order n contains a number n of *arbitrary constants.* This general solution can be specialised to a particular solution by assigninig definite values to these constants.

Example 84.

(a) Family or parabolae $y = Cx^2$ as constant C takes different values.

On a particular curve of the family $\frac{\mathrm{d}y}{\mathrm{d}x} = 2Cx$. By substitutiion, eliminate $C \Rightarrow \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2y}{x}$. This is a geometrical statement about slopes.

Note: 1st order differential equation $\leftrightarrow$ 1 arbitrary constant in general solution.

(b)

$$\left.\begin{aligned} x &= A\sin\omega t + B\cos\omega t \\ \frac{\mathrm{d}x}{\mathrm{d}t} &= A\omega\cos\omega t - B\omega\sin\omega t \\ \frac{\mathrm{d}^2x}{\mathrm{d}t^2} &= -A\omega^2\sin\omega t - B\omega^2\cos\omega t \end{aligned}\right\} \Rightarrow \frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \omega^2 x = 0$$

Note: 2nd order differential equation $\leftrightarrow$ 2 arbitrary constants in general solution.

Of course it's the reverse of this process we normally want to perform in order to get the general solution. We then often need a particular solution — which satisfieis certain other conditions — *boundary* or *initial condition.* These allow us to find the arbitrary constants in the solutions.

2.1.2 First Order Differential Equations

Properties and approaches

There are essentially 4 types we can solve *analytically*:

- ***separable***
- ***homogeneous***
- ***linear***
- ***exact*** (in Chapter "Partial Differentiation and Multivariable Calculus" later)

Let's look at them one by one:

(a) **Separable**

$$\frac{\mathrm{d}y}{\mathrm{d}x} = G(x) \cdot H(y)$$

Solve by rearrangement and integration

$$\int^y \frac{\mathrm{d}y}{H(y)} = \int^x G(x)\mathrm{d}x$$

E.g.

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= xy^2e^{-x} \\ \int \frac{1}{y^2}\mathrm{d}y &= \int xe^{-x}\mathrm{d}x \\ -\frac{1}{y} &= -xe^{-x} - e^{-x} + C \end{aligned}$$

Or singular solution $y = 0$.

If we want the particular solution which passes through $x = 1, y = 1$, then of course we need

$$C = -1 + 2e^{-1} \quad \text{and} \quad \frac{1}{y} = (x+1)e^{-x} + 1 - 2e^{-1}$$

(b) **Homogeneous**

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f\left(\frac{y}{x}\right)$$

Substitution $\frac{y}{x} = u$, i.e. a new dependent variable, and do implicit differentiation on $y = ux$ with respect to x, we get

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= u + x\frac{\mathrm{d}u}{\mathrm{d}x}(= f(u)) \quad \textbf{\textit{(Remember!)}} \\ f(u) - u &= \frac{x\mathrm{d}u}{\mathrm{d}x} \\ \int \frac{\mathrm{d}u}{f(u) - u} &= \int \frac{\mathrm{d}x}{x} \\ &\vdots \end{aligned}$$

E.g.

(i)

$$\begin{aligned} x^2\frac{\mathrm{d}y}{\mathrm{d}x} + xy - y^2 = 0 \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \left(\frac{y}{x}\right)^2 - \frac{y}{x} \\ \frac{\mathrm{d}u}{\mathrm{d}x} &= \frac{u^2 - 2u}{x} \\ &\vdots \end{aligned}$$

(ii)

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x+y-3}{x-y+1}$$

This does not look homogeneous as it stands, but can be made so by substituting $x = 1+X, y = 2+Y$, and the expression becomes

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \frac{X+Y}{X-Y} = \frac{1+\left(\frac{Y}{X}\right)}{1-\left(\frac{Y}{X}\right)}$$

Then let $\frac{Y}{X} = u(X)$,

$$\Rightarrow \int \left(\frac{1-u}{1+u^2}\right) \mathrm{d}u = \int \frac{\mathrm{d}X}{X}$$

Eventually, the equation becomes

$$\tan^{-1}\frac{Y}{X} - \frac{1}{2}\ln\left(1+\frac{Y^2}{X^2}\right) = \ln X + C$$

$$\tan^{-1}\left(\frac{y-2}{x-1}\right) - \frac{1}{2}\ln\left[(x-1)^2 + (y-2)^2\right] = C$$

Note: If we have e.g. $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x+y-3}{2(x+y)-7}$, then substitute $v(x) = x+y$ will work!

(c) **Linear**

$$\frac{\mathrm{d}y}{\mathrm{d}x} + F(x)y = G(x)$$

1st power only for y and $\frac{\mathrm{d}y}{\mathrm{d}x}$. We apply an *integrating factor* $R(x)$:

$$R(x) = \exp\left[\int^x F(x)\mathrm{d}x\right]$$

This allows us to form the expression

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[y\exp\left(\int^x F(x)\mathrm{d}x\right)\right] = G(x)\exp\left(\int^x F(x)\mathrm{d}x\right)$$

and then integrate...

E.g.

$$(x+2)\frac{\mathrm{d}y}{\mathrm{d}x} - 4y = (x+2)^6$$
$$\frac{\mathrm{d}y}{\mathrm{d}x} - \frac{4}{x+2} = (x+2)^5$$
$$\Rightarrow F(x) = -\frac{4}{x+2}, G(x) = (x+2)^5$$

Therefore,

$$R(x) = \exp\left[-\int^x \left(\frac{4}{x+2}\right)\mathrm{d}x\right] = \cdots = K(x+2)^{-4}$$

Subsequently, take $K = 1$ W.L.O.G.:

$$(x+2)^{-4}\frac{\mathrm{d}y}{\mathrm{d}x} - 4(x+2)^{-5}y = \frac{\mathrm{d}}{\mathrm{d}x}\left[y(x+2)^{-4}\right] = x+2$$

As such,

$$y(x+2)^{-4} = \frac{1}{2}x^2 + 2x + C \quad \text{(Put } C \text{ at the right time!)}$$
$$y(x) = \left(\frac{1}{2}x^{2+2x+C}\right)(x+2)^4$$

(So e.g. $y(0) = 8 \Rightarrow C = \frac{1}{2}$)

Novelties!

(i) Bernoulli equation (See Equation(2.4))
A nonlinear equation rendered linear by a substitution $u = y^{1-n}$...

(ii) E.g.

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{x+e^y}$$

It is <u>nonlinear</u> for $y(x)$ but <u>linear</u> for $x(y)$:

$$\frac{\mathrm{d}x}{\mathrm{d}y} - x = e^y \Rightarrow \ldots$$

2.1.3 'Special' Second Order Differential Equations

Definition 85. General Explicit form is

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = F\left(x, y, \frac{\mathrm{d}y}{\mathrm{d}x}\right)$$

(a) $y, \frac{\mathrm{d}y}{\mathrm{d}x}$ **missing**, i.e.

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = f(x)$$

Just integrate twice!

(b) $x, \frac{\mathrm{d}y}{\mathrm{d}x}$ **missing**, i.e.

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = f(y)$$

Warning: Do not write $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{1}{\frac{\mathrm{d}^2x}{\mathrm{d}y^2}}$. However, it may be true, but for what class of functions $y(x)$?

Let $\frac{\mathrm{d}y}{\mathrm{d}x} = p$,

$$\Rightarrow \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\mathrm{d}p}{\mathrm{d}x} = \frac{\mathrm{d}p}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} = p\frac{\mathrm{d}p}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{1}{2}p^2\right)$$

This substitution is effective because it eliminates x, so that the equation becomes separable for p and y.

Then we can integrate $\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{1}{2}p^2\right) = f(y)$ w.r.t. y to get $p(y)$. Then using the definition of p,

$$x = \int \frac{\mathrm{d}y}{p(y)}$$

The same is obtained by multiplying the original equation by $\frac{\mathrm{d}y}{\mathrm{d}x}$ and recognizing $\frac{\mathrm{d}y}{\mathrm{d}x} \cdot \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{2}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]$

Example:

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\omega^2 y$$

with ω being a real constant. (It is a simple harmonic motion.)

$$\Rightarrow \frac{1}{2}p^2 = -\frac{1}{2}\omega^2y^2 + C$$

Let $C = \frac{1}{2}\omega^2\overline{A}^2$. We therefore get

$$\frac{1}{p} = \frac{\mathrm{d}x}{\mathrm{d}y} = \pm\frac{1}{\omega(\overline{A}^2 - y^2)^{\frac{1}{2}}}$$

$$\Rightarrow \omega x + \overline{B} = \pm\sin^{-1}\frac{y}{\overline{A}}$$

$$\begin{aligned} y &= \overline{A}\sin(\omega x + \overline{B}) \text{ W.L.O.G} \\ &= A\sin\omega x + B\cos\omega x \end{aligned}$$

(c) y **missing**, i.e.

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = f\left(x, \frac{\mathrm{d}y}{\mathrm{d}x}\right)$$

We put $\frac{\mathrm{d}y}{\mathrm{d}x} = p$, so

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\mathrm{d}p}{\mathrm{d}x} = f(x, p)$$

i.e. First order $p(x)$. This substitution is effective because it eliminates y, so that the equation becomes <u>separable</u> for p and x.

Solve for $p(x)$ then integrate $\Rightarrow y(x)$.

<u>Example</u>: Radius of curvature

$$\frac{\left[1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{\frac{3}{2}}}{\frac{\mathrm{d}^2y}{\mathrm{d}x^2}} = a \quad (a \text{ is an arbitrary constant})$$

$$\Rightarrow \frac{\mathrm{d}p}{\mathrm{d}x} = \frac{1}{a}(1+p^2)^{\frac{3}{2}}$$

$$\Rightarrow \frac{x}{a} + C = \int \frac{\mathrm{d}p}{(1+p^2)^{\frac{3}{2}}} \quad \text{i.e.} \quad \frac{x}{a} - \frac{A}{a} = \frac{p}{(1+p^2)^{\frac{1}{2}}}$$

$$\Rightarrow \frac{\mathrm{d}y}{\mathrm{d}x} = p = \pm\frac{x - A}{\left[a^2 - (x-A)^2\right]^{\frac{1}{2}}}$$

$$\Rightarrow y = B \mp \left[a^2 - (x-A)^2\right]^{\frac{1}{2}} \quad \text{i.e.} \quad (x-A)^2 + (y-B)^2 = a^2$$

So they are all circles of radius a!

(d) x **missing**, i.e.

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = f\left(y, \frac{\mathrm{d}y}{\mathrm{d}x}\right)$$

Yet again, let $\frac{\mathrm{d}y}{\mathrm{d}x} = p$, so

$$p\frac{\mathrm{d}p}{\mathrm{d}y} = f(y, p)$$

i.e. First order $p(y)$. So we solve for $p(y)$, then find $x = \int \frac{\mathrm{d}y}{p(y)}$.

Example:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = -\omega^2 y \mp 2k\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2$$

SHM with resistance proportional to $(\text{speed})^2$.
Hint: Solving this equation is the perfect application for solving Bernoulli Equation!

(e) **Linear Equations**, i.e. $y, \frac{\mathrm{d}y}{\mathrm{d}x}$ only occur to 1st power, if at all. So no products of y and $\frac{\mathrm{d}y}{\mathrm{d}x}$. The following section is dedicated to explaining the approach to solve linear differential equations.

General case — Linear Equations

The general form is, for order n,

$$\begin{aligned} \mathscr{L}y = a_0(x)\frac{\mathrm{d}^n y}{\mathrm{d}x^n} + a_1(x)\frac{\mathrm{d}^{n-1} y}{\mathrm{d}x^{n-1}} + a_2(x)\frac{\mathrm{d}^{n-2} y}{\mathrm{d}x^{n-2}} + \cdots \\ + a_{n-1}(x)\frac{\mathrm{d}y}{\mathrm{d}x} + a_n(x)y = f(x) \end{aligned} \tag{2.7}$$

where $a_0, a_1, \ldots, a_n$ and $f(x)$ are known functions of x only.
$\mathscr{L}$ is a **linear operator**, operating on $y(x)$:

$$\mathscr{L} \equiv \left[a_0\frac{\mathrm{d}^n}{\mathrm{d}x^n} + a_1\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} + \cdots + a_n\right]$$

The equation(2.7) is called ***homogeneous*** iff $f(x) = 0$ and ***inhomogeneous*** iff $f(x) \neq 0$.
The homogeneous equation $\mathscr{L}y = 0$ has n independent solutions $y_1(x), y_2(x)$,

$\ldots, y_n(x)$ apart from *trivial* $y(x) = 0$. That is to say that $\mathscr{L} y_i(x) = 0$ for $i = 1, 2, \ldots, n$. (***Independence*** is an algebraic property...) Because of the linearity of $y_i(x)$ we find that the most general solution of the homogeneous equation $\mathscr{L} y = 0$ is given by

$$y(x) = A_1 y_1(x) + A_2 y_2(x) + \cdots + A_n y_n(x) \tag{2.8}$$

with $A_1, A_2, \ldots, A_n$ being arbitrary constants. This is because

$$\mathscr{L} y = \mathscr{L}\left(\sum_{i=1}^{n} A_i y_i(x)\right) = \sum_{i=1}^{n} A_i(\mathscr{L} y_i(x)) = 0$$

Of course equation(2.8) contains n arbitrary constants in accord with the order n of the differential equation.

For the inhomogeneous equation ($\mathscr{L} y = f(x)$(2.7)), the expression(2.8) is called the ***complementary functions*** (CF) of equation(2.7). Any solution of the inhomogeneous equation(2.7), say $Y(x)$, is called a ***particular integral*** (PI) of equation(2.7). The most general solution of equation(2.7) is thus

$$y(x) = (\text{CF}) + (\text{PI})$$

This contains n arbitrary constants as required/expected!

The constants can be specified in practice to produce a particular solution which satisfies (n) initial/boundary conditions.
Note

(a) For any two solutions $Y_1(x), Y_2(x)$ of equation(2.7), their difference satisfies

$$\mathscr{L}(Y_1 - Y_2) = \mathscr{L} Y_1 - \mathscr{L} Y_2 = f(x) - f(x) = 0$$

(b) Generally, finding $y_1(x), y_2(x), \ldots, y_n(x)$ functions might be very tough — our differential equation has generally variable coefficients after all! So we look at the most common case we need to study — constant coefficients! W.L.O.G.:

$$a_0(x) = 1, a_1(x) = a_1, a_2(x) = a_2, \ldots, a_n(x) = a_n$$

Linear Equations — Second Order, Constant Coefficients

Consider

$$\mathscr{L}y = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + a_1 \frac{\mathrm{d}y}{\mathrm{d}x} + a_2 y = f(x) \tag{2.9}$$

Alternatively, in terms of notation,

$$\mathscr{L}y = y'' + a_1 y' + a_2 y = f(x)$$

Overall flow of solving the equation is to firstly find CF then PI,

$$\Rightarrow y(x) = \mathrm{CF} + \mathrm{PI}$$

Finding the CF We need to solve

$$\mathscr{L}y = \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + a_1 \frac{\mathrm{d}y}{\mathrm{d}x} + a_2 y = 0 \tag{2.10}$$

Try a solution of the form $y = e^{\lambda x}$ where λ is a constant — which we need to find! (It works by demonstration.) Evidently,

$$(\lambda^2 + a_1\lambda + a_2)e^{\lambda x} = 0$$

The exponential cannot help — for any λ let alone for all x. So

$$\lambda^2 + a_1\lambda + a_2 = 0 \tag{2.11}$$

as the auxiliary equations. In general, there are two distinct roots λ_1, λ_2 of this quadratic, so that $e^{\lambda_1 x}, e^{\lambda_2 x}$ are solutions of equation(2.10), i.e.

$$\mathscr{L}\left(e^{\lambda_1 x}\right) = 0 = \mathscr{L}\left(e^{\lambda_2 x}\right)$$

Because of the linearity property of $\mathscr{L}$ we have

$$y_{\mathrm{CF}} = A_1 e^{\lambda_1 x} + A_2 e^{\lambda_2 x}$$

where A_1, A_2 are two arbitrary constants and $\mathscr{L}y_{\mathrm{CF}} = 0$ as required.

If the roots of (2.11) are equal, i.e. $\lambda_1 = \lambda_2 = \lambda$, then certainly $A_1 e^{\lambda x}$ is a solution of (2.10) with *one* arbitrary constant — we need *another*! A second linearly independent solution is given by $A_2 x e^{\lambda x}$, so that we have

$$y_{\mathrm{CF}} = A_1 e^{\lambda x} + A_2 x e^{\lambda x}$$

We can see this easily: (2.11) must take the form $(\lambda + \frac{a_1}{2})^2 = 0$ since $a_2 = \frac{a_1^2}{4}$ and $\lambda = -\frac{a_1}{2}$ (repeated root). Then substituting $xe^{\lambda x}$ into (2.10) we have

$$\mathscr{L}\left(xe^{\lambda x}\right) = (2\lambda + a_1)e^{\lambda x} + (\lambda^2 + a_1\lambda + a_2)xe^{\lambda x} = 0$$

as required. Here, n in $\mathscr{L}$ is 2.

Example 86.

1. $$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 5\frac{\mathrm{d}y}{\mathrm{d}x} + 6y = 0$$
   $\Rightarrow \lambda^2 + 5\lambda + 6 = 0,\ \lambda = -3, -2.$ So
   $$y(x) = A_1e^{-3x} + A_2e^{-2x}$$

2. $$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 4\frac{\mathrm{d}y}{\mathrm{d}x} + 4y = 0$$
   $\Rightarrow \lambda^2 + 4\lambda + 4 = 0,\ \lambda = -2, -2.$ So
   $$y(x) = A_1e^{-2x} + A_2xe^{-2x}$$

What about *complex roots* of (2.11)? (assuming $a_1, a_2 \in \mathbb{R}$) We know that the roots are complex conjugates, i.e. $\lambda_{1,2} = \alpha \pm i\beta, \alpha, \beta \in \mathbb{R}$. Now, formally our solution is, as above,

$$y = A_1e^{(\alpha+i\beta)x} + A_2e^{(\alpha-i\beta)x}$$

Since $\beta \neq 0$ here since the roots cannot be equal! so we can rewrite in alternative forms:

$$y = e^{\alpha x}\left[A_1e^{i\beta x} + A_2e^{-i\beta x}\right] = e^{\alpha x}\left[C_1\cos\beta x + C_2\sin\beta x\right]$$

where A_1, A_2 or C_1, C_2 can be taken as our arbitrary constants. (Naturally, $C_1 = A_1 + A_2, C_2 = (A_1 - A_2)i$ by De Moivre.)

Example 87.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2k\frac{\mathrm{d}x}{\mathrm{d}t} + \omega^2 x = 0$$

which is the equation for damped harmonic oscillator ($k > 0$).

$$\lambda^2 + 2k\lambda + \omega^2 = 0, \quad \lambda_{1,2} = -k \pm \sqrt{k^2 - \omega^2}$$

and

$$x(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}$$

in general. This can be broken down into different cases.

(1) $k = 0$, i.e. *No Damping.*

$$x = A_1 e^{i\omega t} + A_2 e^{-i\omega t} = C_1 \cos\omega t + C_2 \sin\omega t$$

(2) $k^2 < \omega^2$, i.e. *Light Damping.*

$$x = A_1 e^{-kt+i\omega t} + A_2 e^{-kt-i\omega t} = (C_1 \cos\omega t + C_2 \sin\omega t)e^{-kt}$$

with $\omega = (\omega^2 + k^2)^{\frac{1}{2}}$.

(3) $k^2 > \omega^2$, i.e. *Heavy Damping.*

$$x = A_1 e^{-|\lambda_1|t} + A_2 e^{-|\lambda_2|t}$$

since λ_1, λ_2 are each neagative real.

(4) $k^2 = \omega^2$, i.e. *Critical Damping.*

$$\lambda_1 = \lambda_2 = -k \Rightarrow x = (A_1 + A_2 t)e^{-kt}$$

Note: $x(t)$ behaviours for various cases!

Finding a PI Now we have the CF we need any particular solution of (2.9), in order to complete the job of finding the general solution. The PI is *not* unique! Our guide is the form of the function $f(x)$ on RHS.

(a) *polynomial in x*
Try a polynomial for the PI and choose the coefficients to fit! Example:

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} - 3\frac{\mathrm{d}y}{\mathrm{d}x} + 2y = x$$

Try PI $= ax^2 + bx + c$, where we need to find a, b, c. This method is often known as the method of undetermined coefficients.
We now determine them! (SIAS — Suck It And See)

$$2a - 3(2ax + b) + 2(ax^2 + bx + c) = x$$

By comparing the coefficients, we can obtain

$$a = 0, b = \frac{1}{2}, c = \frac{3}{4} \Rightarrow y_{\mathrm{PI}} = \frac{1}{2}x + \frac{3}{4}$$

Since $y_{\mathrm{CF}} = A_1e^x + A_2e^{2x}$ for this equation, then the general solution can be written as

$$y(x) = A_1e^x + A_2e^{2x} + \frac{1}{2}x + \frac{3}{4}$$

Note: Our inclusion of ax^2 term in our trial PI has been self-correcting since it emerged that $a = 0$. This is always so; the method gives what is needed!

(b) *multiple of e^{bx}*
The obvious choice for the PI is Ae^{bx}, since the linear operator $\mathscr{L}$ generates only terms of this type — choose A to fit! But there are two cases to consider:

(i) e^{bx} *not* in y_{CF}, i.e. $\mathscr{L}(e^{bx}) \neq 0$
Example:

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 5\frac{\mathrm{d}y}{\mathrm{d}x} + 6y = 7e^{8x}$$

with

$$y_{\mathrm{CF}} = A_1e^{-3x} + A_2e^{-2x}$$

Try $y_{\mathrm{PI}} = Ae^{8x}$, then

$$Ae^{8x}[64 + 40 + 6] = 7e^{8x} \Rightarrow A = \frac{7}{110}$$

and general solution is

$$y(x) = y_{\text{CF}} + \frac{7}{110}e^{8x}$$

(ii) e^{bx} is *contained* in y_{CF}, i.e. $\mathscr{L}e^{bx} = 0$

Our trial solution in (i) now does not work! We might hope (anticipate) that xe^{bx} might be involved, and just try it... (SIAS)

A more 'automatic' approach is to take the Ae^{bx} from the CF (where A was constant) and try a PI of the form $A(x)e^{bx}$ — called ***variation of parameters***. We expect that $A(x)$ will be a polynomial in x!

Example:

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 3x + 2y = e^{-x}$$

with

$$y_{\text{CF}} = A_1e^{-x} + A_2e^{-2x}$$

Try $y_{PI} = A(x)e^{-x}$.

$$\Rightarrow (A'' - 2A' + A)e^{-x} + 3(A' - A)e^{-x} + 2Ae^{-x} = e^{-x}$$

By comparing the coefficients, we get

$$A'' + A' = 1$$

Afterwards, integrate with respect to x once and we get

$$A' + A = x + \overline{C_1}$$

Solving this first-order linear equation, and we get

$$A = x + C_1 + C_2e^{-x}$$

$$\Rightarrow y_{\text{PI}} = A(x)e^{-x} = xe^{-x} + C_1e^{-x} + C_2e^{-2x}$$

Take PI $= xe^{-x}$ (W.L.O.G), we can obtain

$$y(x) = A_1e^{-x} + A_2e^{-2x} + xe^{-x}$$

Of course if the auxiliary equation has equal roots then y_{CF} has xe^{bx} too! However the variabtion of parameters still works — or alternatively (a trial polynomial)(e^{bx}).

Example:

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 4\frac{\mathrm{d}y}{\mathrm{d}x} + 4y = e^{-2x}$$

with

$$y_{\mathrm{CF}} = A_1e^{-2x} + A_2xe^{-2x}$$

We can then set PI as

$$y_{\mathrm{PI}} = A(x)e^{-2x} \Rightarrow \ldots A'' = 1 \Rightarrow A = \frac{x^2}{2} + [\overline{A_1} + \overline{A_2}x]$$

$$\Rightarrow y(x) = A_1e^{-2x} + A_2xe^{-2x} + \frac{x^2}{2}e^{-2x}$$

(c) e^{bx} is *polynomial* in x

Try PI $= C(x)e^{bx}$ where $C(x)$ is a polynomial with coefficients to be found — as in (a), (b) above.

(d) sines, cosines, sinh, cosh

We *either* just recognize the pattern and put e.g. $A\cos() + B\sin()$ *or* $A\cosh() + B\sinh()$, etc.

OR

Make use of exponentials — maybe complex ones using $e^{ix} = \cos x + i\sin x$, etc.

Example:

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + 3\frac{\mathrm{d}y}{\mathrm{d}x} + 2y = e^x\cos x$$

with

$$y_{\mathrm{CF}} = A_1e^{-x} + A_2e^{-2x}.$$

There is no obvious trouble with this CF...

(1) Try $y_{\mathrm{PI}} = Be^x\cos x + Ce^x\sin x$ because $\mathscr{L}(y_{\mathrm{PI}})$ produces terms of a similar type. Substitute in and equate coefficients of $e^x\cos x, e^x\sin x$ on the two sides $\Rightarrow B = \frac{1}{10}, C = \frac{1}{10}$.

OR

(2) Put RHS $= \frac{1}{2}e^{(1+i)x} + \frac{1}{2}e^{(1-i)x} (= \Re\left(e^{(1+i)x}\right))$. Then try

$$y_{\text{PI}} = C_1 e^{(1+i)x} \Rightarrow [(1+i)^2 + 3(1+i) + 2]C_1 = 1$$

and $C_1 = \frac{1}{5(1+i)} = \frac{1}{10}(1-i)$, and

$$y_{\text{PI}} = \Re\left[\frac{1}{10}(1-i)e^{(1+i)x}\right] = \frac{1}{10}e^x \cos x + \frac{1}{10}e^x \sin x$$

Naturally, we might need to be adaptable if we find polynomials on RHS in $f(x)$ as well, or the 'equal roots' case... However something to beware:

<u>Example</u>:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + 3\frac{\mathrm{d}y}{\mathrm{d}x} + 2y = \cosh 2x$$

with

$$y_{\text{CF}} = A_1 e^{-x} + A_2 e^{-2x}$$

If we try $y_{\text{PI}} = C_1 \cosh 2x + C_2 \sinh 2x$, we would find C_1, C_2 not defined...

$$\begin{cases} 6C_1 + 6C_2 = 1 \\ 6C_1 + 6C_2 = 0. \end{cases}$$

Why?! Well $\cosh 2x = \frac{1}{2}(e^{2x} + e^{-2x})$ and one of these exponentials *is* in y_{CF}. The better one is

$$y_{\text{PI}} = \frac{1}{24}e^{2x} - \frac{1}{2}xe^{-2x}$$

using earlier results.

<u>Conclusion</u>: Try to use complex numbers, because it avoids "clashing" with hyperbolic functions, and also prevents calculation mistakes, like what would happen when differentiating sines and cosines.

Of course we might finally need to specialise our general solution to the particular solution that satisfies particular boundary conditions.

<u>Example</u>:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + \frac{\mathrm{d}y}{\mathrm{d}x} - 6y = \sin x + xe^{2x}$$

subject to $y(0) = 0$, $\frac{\mathrm{d}y}{\mathrm{d}x}(0) = 0$. The general solution is

$$y(x) = A_1 e^{-3x} + A_2 e^{2x} - \frac{1}{50}(\cos x + 7\sin x) + \frac{e^{2x}}{50}(5x^2 - 2x)$$

and then

$$\left.\begin{aligned} 0 &= A_1 + A_2 - \frac{1}{50} \\ 0 &= -3A_1 + 2A_2 - \frac{7}{50} - \frac{1}{25} \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} A_1 &= -\frac{7}{250} \\ A_2 &= \frac{12}{250}. \end{aligned}\right.$$

2.1.4 Equations with variable coefficients

Special types to meet later (Bessel, Legendre, etc.) ...
A Novelty due to Euler (+ Cauchy!) If W.L.O.G.

$$x^n \frac{\mathrm{d}^n y}{\mathrm{d}x^n} + b_1 x^{n-1} \frac{\mathrm{d}^{n-1} y}{\mathrm{d}x^{n-1}} + \cdots + b_n y = f(x)$$

with $b_1, b_2, \ldots, b_n$ constants.

(i) $f(x) = 0$. Try $y = x^\lambda \Rightarrow$ n values of λ in general.

$$y(x) = A_1 x^{\lambda_1} + A_2 x^{\lambda_2} + \cdots + A_n x^{\lambda_n}$$

with n arbitrary constants.

(ii) $f(x) \neq 0$. The method in (i) above might not be nice for PI! So put $x = e^t$ to *stretch* the independent variable, becoming a *linear equation for* $y(t)$ which has constant coefficients.

Example:

$$x^2 \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + 3x \frac{\mathrm{d}y}{\mathrm{d}x} + y = x^3.$$

Let $x = e^t$, so $\frac{\mathrm{d}x}{\mathrm{d}t} = e^t = t$,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}} = \frac{1}{e^t}\frac{\mathrm{d}y}{\mathrm{d}t}$$

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}y}{\mathrm{d}x}}{\frac{\mathrm{d}x}{\mathrm{d}t}} = \frac{\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-t}\frac{\mathrm{d}y}{\mathrm{d}t}\right)}{e^t} = -e^{-2t}\frac{\mathrm{d}y}{\mathrm{d}t} + e^{-2t}\frac{\mathrm{d}^2 y}{\mathrm{d}t^2}.$$

The equation therefore becomes

$$\left(\frac{\mathrm{d}^2y}{\mathrm{d}t^2} - \frac{\mathrm{d}y}{\mathrm{d}t}\right) + 3\frac{\mathrm{d}y}{\mathrm{d}t} + y = e^{3t}$$

i.e.

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} + 2\frac{\mathrm{d}y}{\mathrm{d}t} + y = e^{3t}.$$

So

$$y(t) = A_1e^{-t} + A_2te^{-t} + \frac{1}{16}e^{3t}$$

and

$$y(x) = \frac{A_1}{x} + \frac{A_2}{x}\ln x + \frac{1}{16}x^3.$$

We should note that $x > 0$ and $x < 0$ need to be treated separately since $x = 0$ is an evident singularity. For $x < 0$ we would need to substitute $x = -e^t$ in the above method.

2.2 Difference Equations

2.2.1 Definitions and Examples

(Recurrence relations, maps, discrete dynamical systems, . . .) From variables whose change is '*continuous*', we now consider variables which are '*discrete*'. ('Season to season', 'one accounrting period to the next', etc.) We have a *dependent variable* $U(n)$ with *integer independent variable* n — together with a relation connecting $U(n)$ to $U(n+1), U(n+2), \cdots$.

Note:

(i) ***Order*** corresponds to how many succeeding generations are involved.

(ii) ***Difference equation*** is associated with e.g. $A(n+1) - A(n) = f[A(n)]$, for instance.

Example 88.

(a) Fibonacci Sequence

Leonardo of Pisa wondered about how many rabbit pairs would be produced in the nth generation starting from a single pair and supposing that any pair from one generation produces a new pair each generation after an initial gap...

$$\begin{cases} U(n) & = 1 \quad 1 \quad 2 \quad 3 \quad 5 \quad 8 \quad 13 \cdots \\ n & = 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad \ \ 7 \cdots \end{cases}$$

and

$$U(n+2) = U(n) + U(n+1).$$

The equation is homogeneous (because only function of $U(n)$ is present without a single term of $f(n)$), linear, and second order.

(b) Money!

If we have an amount $A(n)$ at the beginning of an accounting period, then the amount at the end of that period (i.e. at the beginning of the next) is

$$A(n+1) = (1 + \frac{R}{100})A(n)$$

where $R\%$ is interest rate. The equation is homogeneous, linear, and first order.

If a payment is made each period, then

$$A(n+1) = (1 + \frac{R}{100})A(n) - P.$$

The equation is inhomogeneous, linear, and first order.

(c) Population Dynamics

Population $P(n)$ of an organism measured in each season is

$$P(n+1) = aP(n) - b[P(n)]^2$$

where a, b are positive. The first term indicates the growth, while the second term indicates the overcrowding or competition. (It is quadratic because it relates to the *interactions* of two entities, and the number of way to choose such as pair from a population is quadratic!)

This is a form of what is known as the ***logistic map***. It is homogeneous, nonlinear, and first order. This turns out to have many different behaviours to that of logistic differential equation.

2.2.2 Linear Difference Equations

Broadly we use methods very similar to those we employed for linear differential equations — particularly terminologies like 'Complementary Function' and 'Particular integral', 'number of arbitrary constants', 'order', ...

Example 89.

(a) Fibonacci Sequence

$$U(n+2) = U(n) + U(n+1)$$

Try $U(n) = A\lambda^n$, where A is an arbitrary constant and λ is a particular constant (to be found). We can therefore obtain the *characteristic equation* (as compared with the *auxiliary equation* in differential equations):

$$\lambda^2 - \lambda - 1 = 0.$$

$$\Rightarrow \lambda_{1,2} = \frac{1}{2} \pm \frac{1}{2}\sqrt{5} = \tau, -\frac{1}{\tau}$$

with $\tau = 1.6180\ldots$, which is the golden number. We therefore get

$$\begin{aligned} U(n) &= A_1\lambda_1^n + A_2\lambda_2^n \\ &= A_1\tau^n + A_2\left(-\frac{1}{\tau}\right)^n. \end{aligned}$$

Substitute in $U(1) = 1, U(2) = 1$, we obtain $A_1 = \frac{1}{\sqrt{5}}, A_2 = -\frac{1}{\sqrt{5}}$.

$$\Rightarrow U(n) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

which is known as the "Binet formula".

A particular interesting identity as an application of the Fibonacci Sequence is the "Cassini's identity":

$$U(n+2)U(n) - [U(n+1)]^2 = (-1)^{n+1}$$

which can show that $13 \times 5 - 8^2 = 1$.

There are other sequences, such as the Lucas sequence, where $U(1) = 1, U(2) = 3$, etc.

(b) MoneyA

$$A(n+1) - \left(1 + \frac{R}{100}\right) A(n) = -P$$

$A(n)_{\mathrm{CF}}$ is obtained by solving LHS = 0. Try

$$A(n) = A\lambda^n \Rightarrow \lambda = 1 + \frac{R}{100}$$

and

$$A(n)_{\mathrm{CF}} = A\left(1 + \frac{R}{100}\right)^n.$$

$$A(n)_{\mathrm{PI}} = C, \text{ where } C = \frac{-P}{1 - (1 + \frac{R}{100})}$$

(The power terms cancel out each other due to the coefficient of $A(n)$. Therefore we only take the coefficient of $A(n)$ and $A(n+1)$.) And so

$$A(n) = A\left(1 + \frac{R}{100}\right)^n - \frac{P}{\frac{-R}{100}}$$

We also need to choose appropriate A so that initial balance is $A(0)$.

Note: The methods employed in the previous exmaples are just like those we used for differential equations which have the property of linearity.

General Case with constant coefficients

$$\begin{aligned}\mathscr{L}U(n) = a_0U(n+m) + a_1U(n+m-1) + a_2U(n+m-2) + \cdots \\ + a_{m-1}U(n+1) + a_mU(n) = f(n)\end{aligned}$$

with $a_0, a_1, \ldots, a_m$ constants. The equation is linear, order m. It is homogeneous iff $f(n) = 0$, and inhomogeneous iff $f(n) \neq 0$.

The General Solution (GS) can always be written as

$$U_{\mathrm{GS}} = U_{\mathrm{CF}} + U_{\mathrm{PI}}$$

where $\mathscr{L}U_{\mathrm{CF}} = 0$, $\mathscr{L}U_{\mathrm{PI}} = f(n)$. U_{CF} has m arbitrary constants, while U_{PI} is any solution i.e. it is not unique.

For the CF with a constant coefficient equation we try $U(n)_{\mathrm{CF}} \propto \lambda^n$

$$\Rightarrow \lambda^n[a_0\lambda^m + a_1\lambda^{m-1} + \cdots + a_{m-1}\lambda + a_m] = 0$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are roots of this characteristic equation. Then

$$U(n)_{\text{CF}} = A_1\lambda_1^n + A_2\lambda_2^n + \cdots + A_m\lambda_m^n$$

with $A_1, A_2, \ldots, A_m$ being arbitrary constants.

Example 90.

(1)

$$U(n+2) + 7U(n+1) - 18U(n) = 0$$
$$\Rightarrow \lambda^2 + 7\lambda - 18 = 0, \lambda_1 = -9, \lambda_2 = 2.$$
$$\Rightarrow U(n) = A_1(-9)^n + A_2(2)^n.$$

What about the equal roots case?

(2)

$$U(n+2) - 6U(n+1) + 9U(n) = 0$$
$$\Rightarrow \lambda^2 - 6\lambda + 9 = 0, \lambda_1 = \lambda_2 = 3.$$

Certainly we have $A_1(3)^n$, but we need something else! — It is $A_2 n(3)^n$.

$$\Rightarrow U(n) = A_1 3^n + A_2 n 3^n.$$

What about a PI? Well, as for differential equations, it all depends on $f(n)$!

(a) $f(n) = Cp^n$ where $p \neq \lambda_1$ or λ_2, and C is a constant.

This is easy! $U(n)_{\text{PI}} = Ap^n$ with A chosen suitably. From our earlier example, we put

$$U(n+2) + 7U(n+1) - 18U(n) = 6(4)^n.$$

Since $4 \neq -9$ or 2 we can write $U(n)_{\text{PI}} = A(4^n)$,

$$A(4^{n+2}) + 7A(4^{n+1}) - 18A(4^n) = 6(4^n)$$

i.e. $16A + 28A - 18A = 6 \Rightarrow A = \frac{3}{13}$. So

$$U_{\text{GS}} = A_1(-9)^n + A_2(2)^n + \frac{3}{13}(4)^n.$$

(b) $f(n) = Cp^n$ where $p = \lambda_1$ (say)

Just as for a differential equations we need a more complicated $U(n)_{\text{PI}} = A(n)\lambda_1^n$, where $A(n)$ is a polynomial in n. Again from out earlier example, we put

$$U(n+2) + 7U(n+1) - 18U(n) = 3(2)^n.$$

Let's say

$$U(n)_{\text{PI}} = A(n)(2)^n = (a + bn + cn^2)(2^n)$$

Well, apparently $a = 0$, after comparing with $U(n)_{\text{CF}}$. Then

$$\begin{aligned}[b(n+2) + c(n+2)^2]2^{n+2} + 7[b(n+1) + c(n+1)^2]2^{n+1} \\ - 18(bn + cn^2)2^n = 3(2^n)\end{aligned}$$

Cancel a factor of 2^n, then the n^2 terms are cancelled, and n terms leave $4(b+4c) + 14(b+2c) - 18b = 0$, and constant terms leave $4(2b + 4c) + 14(b+c) = 3$.

$$\Rightarrow c = 0 \text{ and } b = \frac{3}{22}.$$

So

$$U_{\text{GS}} = A_1(-9)^n + A_2(2)^n + \frac{3}{22}n(2^n)$$

and so on...

Since our equation is linear, we can just add terms together to construct $U(n)_{\text{PI}}$ for quite complicated $f(n)$ on RHS.

Some results can seem very strange! The Binet formula for Fibonacci numbers involved irrational numbers as building blocks — but produced integers!

Example:

$$U(n+2) - 2U(n+1) + 5U(n) = 0$$

with say $U(1) = 6, U(2) = 2$ (so that $U(0) = 2$) which obviously produces a sequence of integers. However,

$$\lambda^2 - 2\lambda + 5 = 0 \Rightarrow \lambda_1 = 1 + 2i, \lambda_2 = 1 - 2i.$$

So

$$U(n) = A_1(1+2i)^n + A_2(1-2i)^n$$

Substitute $n = 0, 1$ into the equation, and we get

$$A_1 = 1 - i, A_2 = 1 + i$$

and

$$U(n) = (1 - i)(1 + 2i)^n + (1 + i)(1 - 2i)^n.$$

So $U(3) = -26$, etc.

(c) $f(n)$ is a polynomial in n

Well here we just need to choose a suitable polynomial and choose the coefficients to fit the case.

Example: Try to find

$$S(n) = 1^2 + 2^2 + \cdots + n^2 = \sum_{r=1}^{n} r^2.$$

If we knew the answer or could guess, then we could confirm using induction. If not we can just recognize that

$$S(n+1) - S(n) = (n+1)^2$$

We can easily see that $\lambda = 1$, implying that

$$S(n)_{\text{CF}} = A(1)^n = A.$$

Then

$$S(n)_{\text{PI}} = an^3 + bn^2 + cn.$$

(Do not need a constant term here since it is already in CF.) So

$$a(n+1)^3 + b(n+1)^2 + c(n+1) - an^3 - bn^2 - cn = (n+1)^2$$

Comparing the coefficients, we get $a = \frac{1}{3}, b = \frac{1}{2}, c = \frac{1}{6}$. So

$$S(n)_{\text{GS}} = A + \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$$

and $A = 0$ since we know $S(0) = 0, S(1) = 1$, etc. So

$$S(n) = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n = \frac{1}{6}n(n+1)(2n+1).$$

This method is constructive, and we can extend the idea to find $\sum_{r=1}^{n} r^3 = \left[\frac{1}{2}n(n+1)\right]^2$, etc.

As always, if we tried a polynomial PI which is too simple, or too complicated, the calculation is self-correcting!

(d) $f(n) = (\text{polynomial in } n)(p)^n$

Just like our previous cases our expectation is

$$U(n)_{\mathrm{PI}} = (\text{suitable polynomial})(p)^n.$$

Then following similar step: matching coefficients, substitute in values, obtain value of the constant if boundary condition is provided, etc.

2.2.3 Differencing and Difference Tables

Definition 91. The (forward) ***difference operator*** Δ is defined by

$$\Delta U(n) = U(n+1) - U(n)$$

so that

$$\begin{aligned}\Delta^2 U(n) &= \Delta[U(n+1) - U(n)] \\ &= \Delta U(n+1) - \Delta U(n) \\ &= [U(n+2) - U(n+1)] - [U(n+1) - U(n)] \\ &= U(n+2) - 2U(n+1) + U(n)\end{aligned}$$

(Attention: binomial coefficients appear in the above process! and this continues on!) Now we can see that $\Delta n^k = (n+1)^k - n^k = kn^{k-1} + \cdots + 1$, and this means that

$$\Delta(\text{polynomial in } n \text{ of degree } k) = (\text{polynomial in } n \text{ of degree } (k-1))$$

We can continue this process of course, $\Delta(\Delta(\Delta(\ldots))) = \Delta^k()$.

$$\Rightarrow \Delta^k(\text{polynomial of degree } k) = (\text{polynomial of degree } 0)$$

and $\Delta^{k+1}(\text{polynomial of degree } k) = 0$.

Note: Successive differencing is a *discrete* analogy to differentiation. Do a comparison with the definition of differentiation at a point. ($\frac{\mathrm{d}^4}{\mathrm{d}x^4}(x^4) = 24$

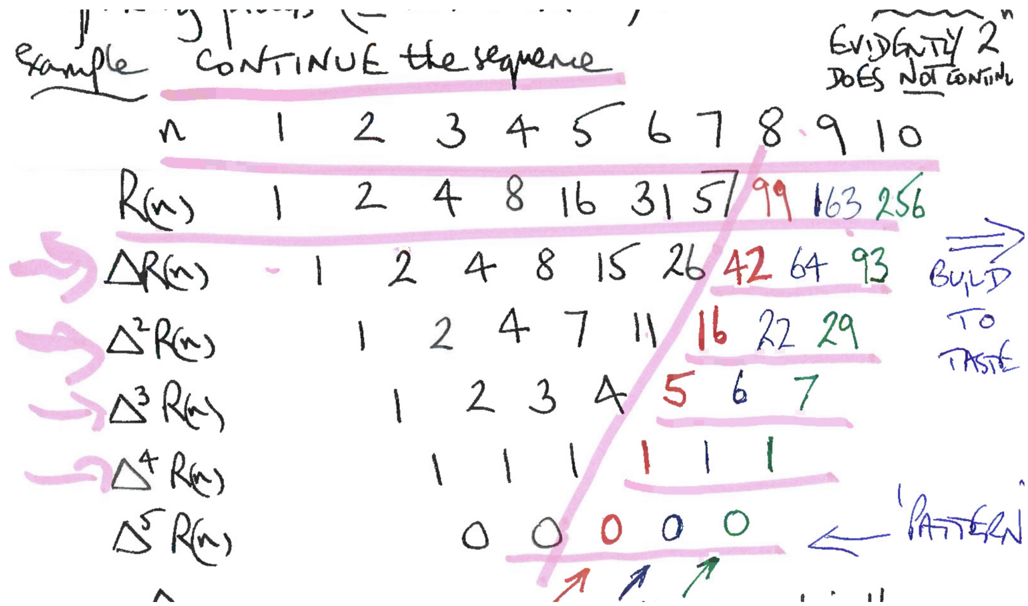


Figure 2.1: graph of reverse differencing porcess

of course!) We can consider the reverse (*inverse*) of the differencing process ($\approx$ integration).

Example 92.

$$
\begin{aligned}
\Delta n^4 &= (n+1)^4 - n^4 = 4n^3 + 6n^2 + 4n + 1 \\
\Delta^2(n^4) &= \Delta(4n^3) + \Delta(6n^2) + \Delta(4n) + \Delta(1) = 12n^2 + 24n + 14 \\
\Delta^3(n^4) &= 24n + 36 \\
\Delta^4(n^4) &= 24 \\
\Delta^5(n^4) &= 0.
\end{aligned}
$$

And and example of the *inverse* process is as shown in figure 2.1. To find out the sequence of $R(n)$ beyond $n = 7$, one can keep on differencing the sequence (which is *polynomial-like*) until its fourth and fifth order, realizing the repetitive 0s and 1s pattern, construct further 1s and 0s, and do the inverse back until order 0, i.e. constructing $R(n)$. The pattern continues, in fact, only when $R(n)$ is a $k = 4$ degree polynomial in n.

<u>Note</u>: (Not in syllabus) There is a discrete analogy to Taylor's expansion, involving Newton's forward difference interpolation formula ...

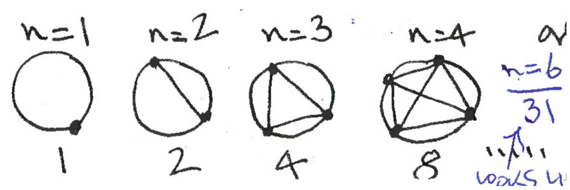


Figure 2.2: Circle Division

The sequence in the inverse process is actually

$$\begin{aligned}
&1 + (n-1) + \frac{1}{2}(n-1)(n-2) + \frac{1}{6}(n-1)(n-2)(n-3) \\
&\qquad\qquad\qquad\qquad + \frac{1}{24}(n-1)(n-2)(n-3)(n-4) \\
&= \binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \binom{n-1}{3} + \binom{n-1}{4} \\
&= \binom{n}{0} + \binom{n}{2} + \binom{n}{4}.
\end{aligned}$$

This expression represents the numbers of distinct regions into which the interior of a circle is partitioned when n distinct boundary points are connected by straight lines, as shown in figure 2.2. This is, however, not easy to prove!

2.2.4 First Order Recurrence/Discrete Nonlinear Systems

Consider $x_{n+1} = F(x_n)$ where $x_n = x(n), x_n \neq 0$. And we have initial choice x_0:

$$\Rightarrow x_1 = F(x_0) \Rightarrow x_2 = F(x_1) = F(F(x_0)) = F^{(2)}(x_0) \Rightarrow \ \ldots$$

This process is called ***iteration*** — some function is used repeatedly — *iterative process*. We can repersent this process graphically, as shown in Figure 2.3.

There are 2 fixed points P_1 and P_2, for which the x values satisfy

$$X = F(X) \Rightarrow X_1, X_2.$$

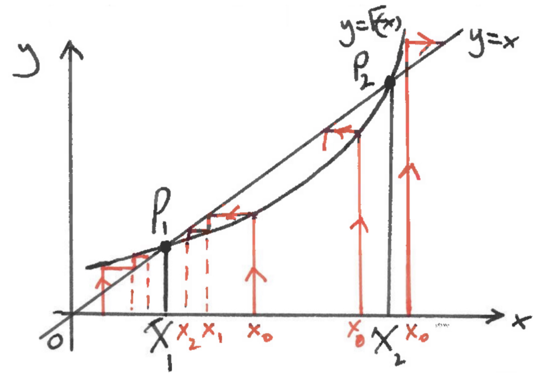


Figure 2.3: 'Cobweb' Diagram

However, the *character* of P_1 and P_2 is very different — initial values x_0 which start near X_1 have x_n which approaches X_1, while those x_0 which start near X_2 certainly are *not* giving x_n which approaches X_2!

Definition 93. X_1 corresponding to P_1 is said to be ***asymptotically stable*** or *attracting*, and is called ***attractor***; X_2 corresponding to P_2 is said to be ***unstable*** or *repelling*, and is called ***repeller***.

How can we distinguish them analytically?

Suppose $x_{n+1} = F(x_n)$ and $X = F(X)$. We put $X = x_n + \epsilon_n$ and imagine x_0 is chosen so that ϵ_0 is 'small' i.e. x_0 is 'near' to X. Let's see how ϵ_n develops (whether x_n converges or diverges to X):

$$X - \epsilon_{n+1} = x_{n+1} = F(x_n) = F(X - \epsilon_n) = F(X) - \epsilon_n F'(X) + \frac{1}{2}\epsilon_n^2 F''(X) + \cdots$$

with the last step using taylor expansion, and by cancelling X and $F(X)$, we get

$$\epsilon_{n+1} = \epsilon_n F'(X) - \frac{1}{2}\epsilon_n^2 F''(X) + \cdots .$$

ϵ_{n+1} can therefore be estimated using different values of the various orders of $F(X)$:

- $F'(X) \neq 0 \Rightarrow \epsilon_{n+1} \approx \epsilon_n F'(X) \Rightarrow \epsilon_n \approx \epsilon_0 [F'(X)]^n$.

  This process is called **first order process**. Then if $|F'(X)| < 1$, then $\epsilon_n \to 0$ and X is an attractor. Otherwise if $|F'(X)| > 1$, then ϵ_n

diverges and X is a repeller. However, if $|F'(X)| = 1$ then it depends on the case — nothing is already proven.

- $F'(X) = 0, F''(X) \neq 0 \Rightarrow \epsilon \approx -\frac{1}{2}\epsilon_n^2 F''(X) \Rightarrow \epsilon_{n+1} \propto \epsilon_n^2$.

  This process is called **second order process**. $\forall \epsilon_0$ sufficiently small, we have $\epsilon_n \to 0$, and X is *always* an attractor. (Proof is not provided here.)

  Note that it is *faster* than first order convergence, therefore it is usually preferred to design a process such that it is second order for studying that particular matter for better result.

- $F'(X) = 0, F''(X) = 0, F'''(X) \neq 0 \Rightarrow \epsilon_{n+1} \propto \epsilon_n^3$.

  This process is called **third order process**.

And so on. The *rate* of convergence increases with the order of the process. Third order process and beyond are usually unnecessary, but occasionally they may be required. In practice we hope for second order, but will often settle for first order.

Example 94.

(a)

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{A}{x_n}\right) = F(x_n)$$

which is a method for finding $\sqrt{A}$. For instance, $A = 12, x_0 = 2, \ldots, x_4 = 3.4641$, etc.

The fixed points are $X = \frac{1}{2}\left(X + \frac{A}{X}\right) \to X = \pm\sqrt{A}$. By drawing the Cobweb diagram, we should see that $x_0 > 0 \Rightarrow x_n \to \sqrt{A}, x_0 < 0 \Rightarrow x_n \to \sqrt{A}$.

Next we find out which order the process is:

$$F'(X) = \frac{1}{2}\left(1 - \frac{A}{X^2}\right) = 0$$
$$F''(X) = \frac{A}{X^3} = \pm\frac{1}{\sqrt{A}} \neq 0.$$

So this is a second order process, and $\pm\sqrt{A}$ are attractors with $\epsilon_{n+1} \propto \epsilon_n^2$.

<u>Exercise</u>: Consider $A < 0$?

(b) Solve

$$f(x) = x^2 - 6x + 2 = 0.$$

We can rearrange this in various ways and write it in iterative process:

(i) $x_{n+1} = 6 - \frac{2}{x}$

(ii) $x_{n+1} = \frac{1}{6}x_n^2 + \frac{1}{3}$

(iii) $x_{n+1} = \sqrt{6x_n - 2}$

(iv) $x_{n+1} = x_n - \frac{x_n^2 - 6x_n + 2}{2x - 6} = \frac{x_n^2 - 2}{2x_n - 6}$.

Examining these (see Problem Sheet 3) we find that (iv) is the 'best buy' in that it is the *only* second order process and it is the only one which allows us to obtain both roots and attractors if we choose x_0 suitably.

(c)

$$x_{n+1} = x_n(2 - Ax_n)$$

which is a method for finding a reciprocal *without* division! ($x_n \to \frac{1}{A}$) It is a second order process.

Note: Examples (a), (b)(iv), (c) are examples of what is now called the *Newton(-Raphson) Method* for finding solutions of $f(x) = 0$:

$$x_{n+1} = F(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Such a process is *normally* at least second order (good!) because

$$F'(x) = 1 - \frac{f'(x)}{f'(x)} + \frac{f(x)f''(x)}{(f'(x))^2} = 0$$

and

$$F''(x) = \frac{f''(x)}{f'(x)} \neq 0 \text{ usually.}$$

However, there are some difficulties in implementing the method successfully, including choosing a value near roots, having multiple roots, etc.

(d) modern, practical, surprising... Population Dynamics

Recall the *logistic map equation*:

$$P(n+1) = aP(n) - b(P(n))^2$$

which is a simple mathematical model with very complicated dynammics. Put $x_n = \frac{b}{a}P(n)$, and we get

$$x_{n+1} = ax_n(1 - x_n)$$

with a being the constant. This is the standard form of logistic map.

Although there is no restriction for mathematical interest, the 'physical' interest is in $0 \leq a \leq 4$ so that $[0, 1] \to [0, 1]$. We can easily see that the maximum value that $x_n(1 - x_n)$ can get is $\frac{1}{4}$, therefore having any $a > 4$ would definitely result in $x_{n+1} > 1 \Rightarrow x_{n+2} < 0$, and let alone $a < 0$. We certainly would not want negative population values!

There are evidently two fixed points satisfying

$$X = aX(1 - X) \Rightarrow X = 0 \text{ and } X = 1 - \frac{1}{a}.$$

Which do we get, and when? Take the first order process and analyse with different ranges of a:

$$|F'(X)| = |a(1 - 2X)|.$$

- $0 \leq a < 1$, $a = 0$ is trivial.

  We can deduce that $X = 0$ is an attractor, while $X = 1 - \frac{1}{a}$ is a repeller. This makes sense because it is a linear model made worse by overcrowding.

- $1 < a < 3$.

  We can decude that $X = 0$ is a repeller, while $X = 1 - \frac{1}{a}$ is an attractor. This makes sense because it is an exponential growth stabilised by overcrowding.

  This behaviour is very similar to that of the logistic differential equation — what follows is definitely not so!

- $a > 3$.

  We can deduce that $X = 0$ and $X = 1 - \frac{1}{a}$ are both repellers.

  So what exactly do we get? We get '*period doubling*'.

  Consider $x_{n+2} = F(x_{n+1}) = F(F(x_n)) = F^{(2)}(x_n) = a[ax_n(1 - x_n)][1 - ax_n(1 - x_n)]$. The fixed points of this satisfy

  $$X = a^2 X(1-X)[1 - aX(1-X)] \tag{2.12}$$

  We still have $X = 0, X = 1 - \frac{1}{a}$ of course, (it is a fixed point on a map done twice.) but now we have two new ones, say X_1 and X_2, satisfying

  $$a^2X^2 - a(a+1)X + (a+1) = 0 \tag{2.13}$$

  which is derived from dividing equation (2.12) with the two known solutions. (Or do factorization accordingly.)

  We also know that $X_1 = F(F(X_1)), X_2 = F(F(X_2))$. As such, choosing X_1, and applying the map once, we can see that $F(X_1) = F(F(F(X_1)))$, i.e. both X_1 and $F(X_1)$ are the two fixed points of the map done twice, satisfying the equation (2.13), which is exactly looking for the two roots which are the fixed points of the logistic map done twice, and there are no other such fixed point except for X_2, thus we must have

  $$X_1 = F(X_2), \quad X_2 = F(X_1).$$

  This forms a *flip* or *2-cycle*. (Before becoming 4-cycle, 8-cycle, etc.) This is an attractor when

  $$\left|\frac{\mathrm{d}}{\mathrm{d}x}F(F(x))\right| < 1$$

  $$\Rightarrow |F'(X_1)F'(X_2)| < 1, |a(1-2X_1)a(1-2X_2)| < 1$$

  and using Vieta's theorem to obtain the sum and product of the two roots from equation (2.13), we get

  $$3 < a < 1 + \sqrt{6}$$

  for positive a. For increasingly larger $a > 1 + \sqrt{6}$, we then obtain 4 cycle $\Rightarrow$ 8 cycle $\Rightarrow \ldots \Rightarrow$ arbitrary number of cycle, or *chaotic behaviour*!

Novelty! The stable windows (when it is still in a certain number of cycle instead of being completely random, or rather *chaotic* behaviour) get shorter in geometrical progression at rate $\frac{1}{4.669...}$, where $4.669\ldots$ is the *Feigenbaum constant*. (The first one. There is another one, which is not introduced by Berkshire, and actually beyond the scope of the current study, according to Wikipedia.) For $3.57 < a \leq 4$, 'Chaos' + periodic windows!

For motivation to study this section, please watch the following youtube video:
https://www.youtube.com/watch?v=ovJcsL7vyrk

For studying in detail, please read the following book recommended by our dear lecturer Frank Berkshire:
https://physicaeducator.files.wordpress.com/2018/02/classical-mechanics-by-kibble-and-berkshire.pdf

2.3 Linear Systems of Differential Equations

2.3.1 definitions and examples

Previously we saw some simple examples of systems of differential equations, where there is more than one dependent variable, e.g. the first order system

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = F(x,y) \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = G(x,y). \end{cases} \tag{2.14}$$

A very important class for us to consider is that of *linear systems*:

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = ax+by \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = cx+dy \end{cases} \tag{2.15}$$

where, in general, a, b, c, d could be functions of time — we take them to be *constants* in our discussion. System (2.15) is called ***homogeneous***. If there are further added constants or functions of time on RHS of (2.15) then the system would be called ***inhomogeneous***.
Notes: (2.15) is a ***coupled system*** in general, in that x and y appear on *each* RHS.

Examples

(a) Combat:

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = -ay \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = -bx. \end{cases}$$

Here $a = 0 = d$; b, c are both negative in (2.15).

(b) Romance!

$$\begin{cases} \dfrac{\mathrm{d}r}{\mathrm{d}t} = ar + bj \\ \dfrac{\mathrm{d}j}{\mathrm{d}t} = cr + dj \end{cases}$$

where a, b, c, d can plausibly be positive or negative! ($r(t)$ is Romeo's love/hate for Juliet at time t. Similarly, $j(t)$ is Juliet's love/hate for Romeo at time t.)

(c) Linear Ordinary Differential Equations of higher order

E.g. Our damped harmonic oscillator

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2k\frac{\mathrm{d}x}{\mathrm{d}t} + \omega^2 x = 0$$

can be written in the form

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = y \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = -\omega^2 x - 2ky. \end{cases} \tag{2.16}$$

(d) General nonlinear systems

In general we can find equilibria of (2.14) by solving $F(x, y) = 0 = G(x, y)$. The local behaviour of x, y near these equilibria is that of a local linear System (via Taylor expansion). This analysis allows us to infer the properties of the full nonlinear systems.

How do we solve *linear systems* like (2.15)?

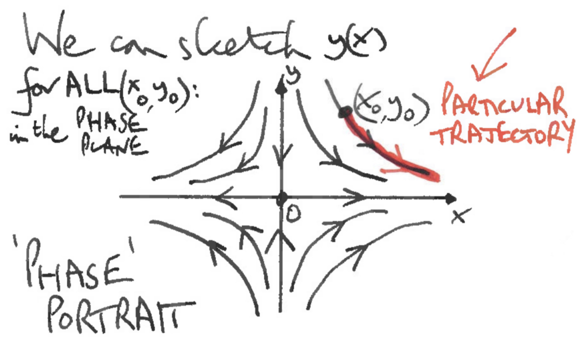


Figure 2.4: 'phase' portrait

A desirable and worthy aim is to try to *decouple* the equations — if necessary by making a suitable change of variables. This is a good idea because if we have a decoupled system, e.g.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 2x, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -3y$$

then for an initial ($t = 0$) point (x_0, y_0) the solution would be $(x, y) = (x_0 e^{2t}, y_0 e^{-3t})$. In this case, as shown in Figure 2.4, all the trajectories are given by $x^3y^2 =$ constant, and the particular solution has $x^3y^2 = x_0^3y_0^2$. Within a family of solutions, the value of x and y changes as the value of t changes, with the direction specified in the diagram.

As such, we can also see that there is one equilibrium point at $O(0, 0)$, where any changes in the value of t does not change the value of x and y. In addition, the equilibrium point is *not stable* since, although having perturbation in the y-axis converges back to 0, perturbations along the x-axis diverges to infinity. So O is definitely not an attractor.

What can be done with a coupled system? e.g.

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = -4x - 3y \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = 2x + 3y \end{cases} \tag{2.17}$$

(For the moment we consider a homogeneous system — some inhomogeneous later.)

Methods

(1) We might recognize that

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+4\right)x=-3y \quad \text{and} \quad \left(\frac{\mathrm{d}}{\mathrm{d}t}-3\right)y=2x$$

$$\Rightarrow \left(\frac{\mathrm{d}}{\mathrm{d}t}-3\right)\left(\frac{\mathrm{d}}{\mathrm{d}t}+4\right)x=-3\left(\frac{\mathrm{d}}{\mathrm{d}t}-3\right)y=-6x$$

$$\Rightarrow \frac{\mathrm{d}^2x}{\mathrm{d}t^2}+\frac{\mathrm{d}x}{\mathrm{d}t}-6x=0.$$

Solve this using the previous methods: $\lambda_1 = 2, \lambda_2 = -3$, so that $x(t) = A_1e^{2t} + A_2e^{-3t}$. Naturally we can then find $y(t)$ from the first rearrangement above:

$$y(t)=-\frac{1}{3}\left(\frac{\mathrm{d}}{\mathrm{d}t}+4\right)x=-2A_1e^{2t}-\frac{1}{3}A_2e^{-3t}$$

and we note that our solution for $x(t), y(t)$ depends on 2 arbitrary constants — as it must!

Afterwards, we can also find $y(x)$ by eliminating t, if we wish. Using the expressions we obtained for $x(t)$ and $y(t)$, we can obtain

$$(x+3y)=-5A_1e^{2t}, \quad (2x+y)=\frac{5}{3}A_2e^{-3t}$$

$$\Rightarrow (x+3y)^3(2x+y)^2=\frac{3125}{9}A_1^3A_2^2.$$

We can then draw the family of trajectories in the (x, y) plane (phase portrait).

(2) We might also note that our (2.17) can be written as

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}}=\frac{2x+3y}{-4x-3y}=\frac{2+3\left(\frac{y}{x}\right)}{-4-3\left(\frac{y}{x}\right)}.$$

This is homogeneous 1st order D.E. We put $\frac{y}{x} = u(x) \Rightarrow x\frac{\mathrm{d}u}{\mathrm{d}x} + u = \frac{2+3u}{-4-3u}$, and then we get

$$x\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{2+7u+3u^2}{-4-3u}$$

solving this equation and we get

$$-\ln x = \frac{3}{5}\ln(1+3u) + \frac{2}{5}\ln(2+u) + C$$

substituting x and y back and eliminating t and we get

$$(x+3y)^3(2x+y)^2 = C.$$

Warning: This method is not favourable as it does not contain any information regarding t— it was got rid of at the very start, therefore no time-depence of x and y, i.e. $x(t), y(t)$. As a result, we cannot do certain things such as drawing the phase portrait!

(3) We might just *notice(!)* that

$$\frac{\mathrm{d}}{\mathrm{d}t}(2x+y) = 2(-4x-3y) + (2x+3y) = -3(2x+y)$$
$$\frac{\mathrm{d}}{\mathrm{d}t}(x+3y) = (-4x-3y) + 3(2x+3y) = 2(x+3y)$$

so that

$$\begin{cases} 2x+y = C_1e^{-3t} \\ x+3y = C_2e^{2t} \end{cases}$$

$$\Rightarrow \begin{cases} x = \dfrac{3}{5}C_1e^{-3t} - \dfrac{1}{5}C_2e^{2t} \\ y = -\dfrac{1}{5}C_1e^{-3t} + \dfrac{2}{5}C_2e^{2t} \end{cases}$$

and of course $(x+3y)^3(2x+y)^2 = C$ again.

All the aforementioned methods have the same phase portrait, as shown in Figure 2.5. The direction of the two straight lines $x+3y=0$ and $2x+y=0$ can be found easily, and the other four trajectories has to follow the same direction as those two straight lines, unless there are two trajectories intersect and have an equilibrium point.

Method (3) gives the germ of a good idea!

(4) How can we arrive at the linear combinations we used previously in an orderly manner and not just 'by inspection' or luck?!

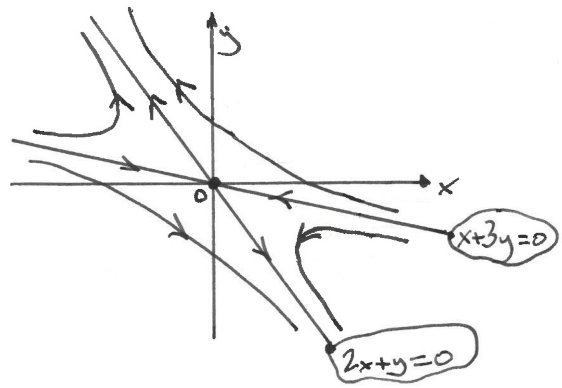


Figure 2.5: phase portrait of (2.17)

We write our system (2.17) in a different way:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -4 & -3 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \tag{2.18}$$

which is $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v} = M\mathbf{v}$ in vector/matrix notation. Now try $\mathbf{v} = \mathbf{V}e^{\lambda t}$ with $\mathbf{V}$ *not* depending on t, i.e. it is a *constant vector*. This implies that

$$M\mathbf{V} = \lambda\mathbf{V}, \text{ i.e. } (M - \lambda I)\mathbf{V} = \mathbf{0}$$

converting to an eigenvalue/eigenvector problem to find appropriate $\lambda, \mathbf{V}$.

For non-trivial $\mathbf{V}$ (i.e. $\neq 0$) then we must have

$$\det\begin{pmatrix} -4-\lambda & -3 \\ 2 & 3-\lambda \end{pmatrix} = 0$$

$\Rightarrow (-4-\lambda)(3-\lambda) + 6 = 0,\ \lambda^2 + \lambda - 6 = 0,\ \lambda_1 = 2, \lambda_2 = -3$. And computing the respective eigenvector, we get

$$\mathbf{V}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \mathbf{V}_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix},$$

or any scalar multiple of these vectors. (Substitute λ_1, λ_2 into $(M - \lambda I)$ and compute its RRE, and read off from RRE.)

We now note the *linearity* of our system, so we can write

$$\begin{pmatrix} x \\ y \end{pmatrix} = B_1 \begin{pmatrix} 1 \\ -2 \end{pmatrix} e^{2t} + B_2 \begin{pmatrix} 3 \\ -1 \end{pmatrix} e^{-3t}$$

where B_1, B_2 are arbitrary constants. Therefore $x = B_1 e^{2t} + 3B_2 e^{-3t}$, $y = -2B_1 e^{2t} - B_2 e^{-3t}$. (Note: results are the same as the previous methods.)

2.3.2 System decoupling

How does the last method do it? From the eigenvectors $\mathbf{V}_1, \mathbf{V}_2$ form the matrix

$$S = (\mathbf{V}_1 \quad \mathbf{V}_2) \qquad \text{i.e. } 2 \times 2$$

and write

$$\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix} = S \begin{pmatrix} \xi \\ \eta \end{pmatrix}.$$

(*Similarity transformation*: tranform linearly from old basis (axis) (x, y) to new basis (axis) (ξ, η). For more background knolwedge on how this works, please refer to subsection 3.6.4.) The vector form of (2.18) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[S \begin{pmatrix} \xi \\ \eta \end{pmatrix}\right] = S \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = MS \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = S^{-1} MS \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

assuming that S is a *nonsingular matrix* (invertible). Then by noting that

$$S^{-1} S = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} (\mathbf{V}_1 \quad \mathbf{V}_2) = I_n$$

where R_1, R_2 are of dimension 1×2. Using $M\mathbf{V} = \lambda \mathbf{V}$, we can deduce that

$$S^{-1} MS = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} MS = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} (\lambda_1 \mathbf{V}_1 \quad \lambda_2 \mathbf{V}_2) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

which is a diagonal matrix. For instance, in (2.18),

$$S = \begin{pmatrix} 1 & 3 \\ -2 & -1 \end{pmatrix}, S^{-1} = \frac{1}{5} \begin{pmatrix} -1 & -3 \\ 2 & 1 \end{pmatrix}$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}2 & 0\\ 0 & -3\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}$$

and this is decoupled! As such, $\frac{\mathrm{d}}{\mathrm{d}t}\xi = 2\xi, \frac{\mathrm{d}}{\mathrm{d}t}\eta = -3\eta \Rightarrow \xi = C_1 e^{2t}, \eta = C_2 e^{-3t}$. Therefore as before,

$$\begin{pmatrix}x\\ y\end{pmatrix} = S\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}C_1 e^{2t} + 3C_2 e^{-3t}\\ -2C_1 e^{2t} - C_2 e^{-3t}\end{pmatrix}$$

General Theory Does this vector/matrix method always work?

For $\mathbf{v} = \begin{pmatrix}x\\ y\end{pmatrix}$, $\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = M\mathbf{v}$, substituting with $\mathbf{v} = \mathbf{V}e^{\lambda t}$, we get $(M - \lambda I)\mathbf{V} = \mathbf{0}$. The solution depends *crucially* on the nature of the eigenvalue/eigenvector problem $\lambda_1, \mathbf{V}_1, \lambda_2, \mathbf{V}_2$.

We note that $\lambda_1 + \lambda_2 =$ *trace* of M (sum of the leading diagonal of M), and $\lambda_1\lambda_2 = \det M$. Both can be derived by looking at the quadratic equation derived from calculating the determinant of $(M - \lambda I)$ to be 0. The equation is *characteristic* for M.

- $\lambda_1 \neq \lambda_2$

  Our example (2.18) was of this type, thereby having $\mathbf{V}_1, \mathbf{V}_2$ eigenvectors. Our system decouples via the *similarity transformation* and $S^{-1}MS$ is a diagonal matrix. This is essentially the situation even when the λ_i are *complex*.

- $\lambda_1 = \lambda_2 = \lambda$

  We've always had some difficulty with the equal roots cases... Here the difficulty presents in that the matrix S, which effects the desired decoupling *may not exist*.

  (i) 2 distinct eigenvectors exist.
  If $\mathbf{V}_1, \mathbf{V}_2$ are distinct for M, λ then it is true that $M\mathbf{v} = \lambda\mathbf{v}$ for any vector, i.e. $M = \lambda I$ and

  $$\begin{pmatrix}x\\ y\end{pmatrix} = B_1\mathbf{V}_1 e^{\lambda t} + B_2\mathbf{V}_2 e^{\lambda t}.$$

  with $\mathbf{V}_1, \mathbf{V}_2$ being two linearly-independent and random vectors in the plane.

(ii) Only one eigenvector exists.

The best that can be done by a similarity transformation is to reduce the system to e.g.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}\lambda & 0\\ 1 & \lambda\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}$$

or equivalent. The diagonal form is *not* achievable. By looking at $\frac{\mathrm{d}\eta}{\mathrm{d}t}$, we realize that

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = \xi + \lambda\eta \Rightarrow \frac{\mathrm{d}\eta}{\mathrm{d}t} - \lambda\eta = \xi,$$

where the CF of η already appears in ξ. So we have to use $te^{\lambda t}$ instead. As such,

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix} = C_1\begin{pmatrix}0\\ 1\end{pmatrix}e^{\lambda t} + C_2\begin{pmatrix}1\\ t\end{pmatrix}e^{\lambda t}.$$

Note: If M is actually a *symmetric matrix* with real entries then λ_1, λ_2 are real, the $\mathbf{V}_1, \mathbf{V}_2$ are real and orthogonal. If we then choose $\mathbf{V}_1, \mathbf{V}_2$ (as we may, if we wish) to be of unit length (normalised), then $S^{-1} = S^T$ and S represents a rotation in the x, y plane (or rotation + reflection). S is an orthogonal matrix.

2.3.3 Typical Phase Portraits

What do the phase portraits look like in the various different cases? The eigenvalues λ_1, λ_2 are evidently crucial in determining the structure and the time t evolution arrows on the trajectories.

The eigenvectors determine the crucial ξ, η decoupled coordinate directions, where appropriate. We note in the interpretation of the diagrams that

$$\xi \propto e^{\lambda_1 t}, \eta \propto e^{\lambda_2 t} \quad \Longrightarrow \quad \xi \propto \eta^{\frac{\lambda_1}{\lambda_2}}, \frac{\mathrm{d}\xi}{\mathrm{d}\eta} \propto \eta^{\frac{\lambda_1}{\lambda_2}-1} \text{ in general.}$$

Thus dominant eigenvector *nearly always* depends on $\left|\frac{\lambda_1}{\lambda_2}\right|$. $\left|\frac{\lambda_1}{\lambda_2}\right| < 1 \Longrightarrow \mathbf{V}_1$ is the dominant eigenvector, and the other trajectories come in *parallel/tangent* to the dominant eigenvector. Now pardon me for the shameleses numerous screenshots of Berkshire's notes for the phase portraits.

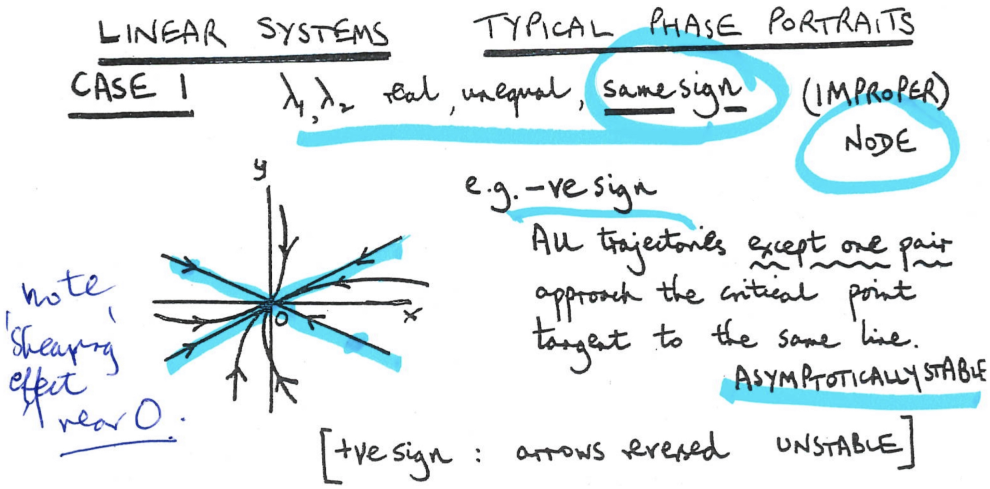


Figure 2.6: case 1

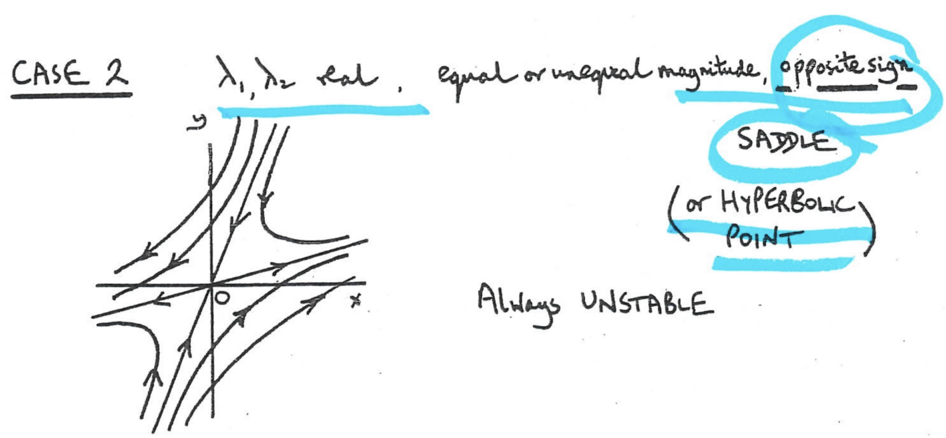


Figure 2.7: case 2

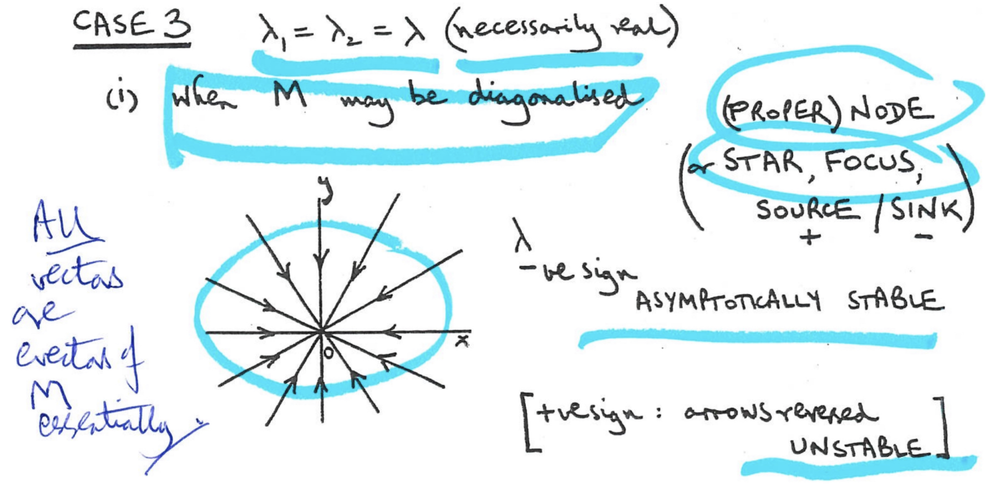


Figure 2.8: case 3 (i)

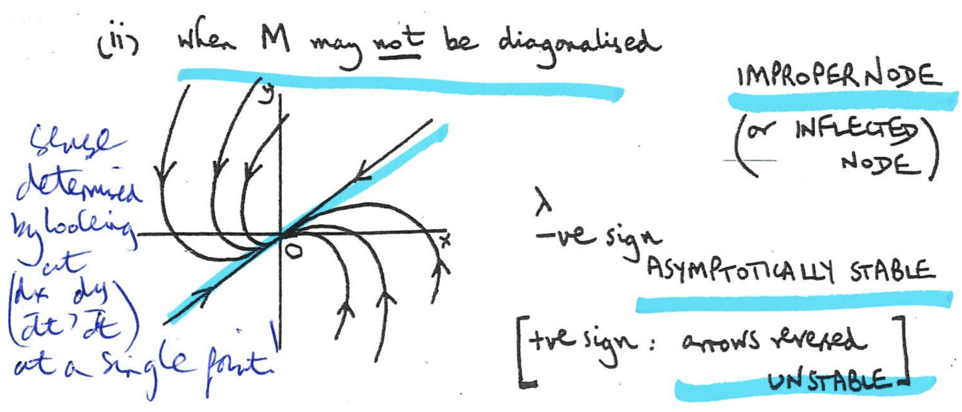


Figure 2.9: case 3 (ii)

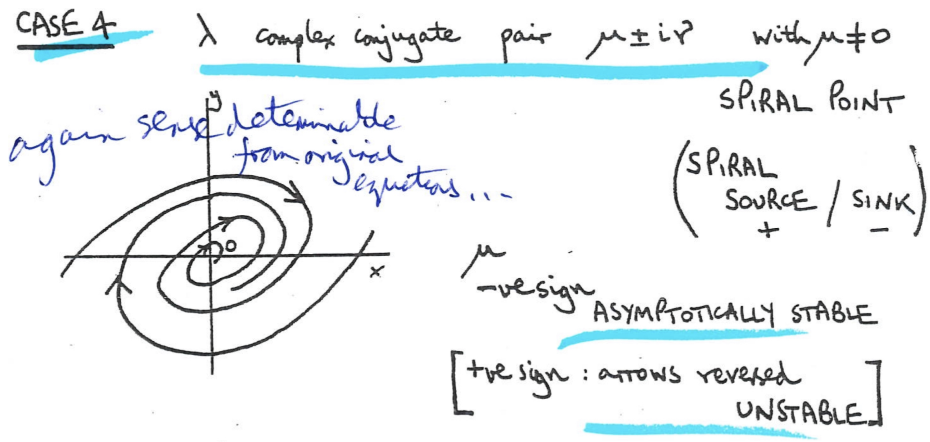


Figure 2.10: case 4

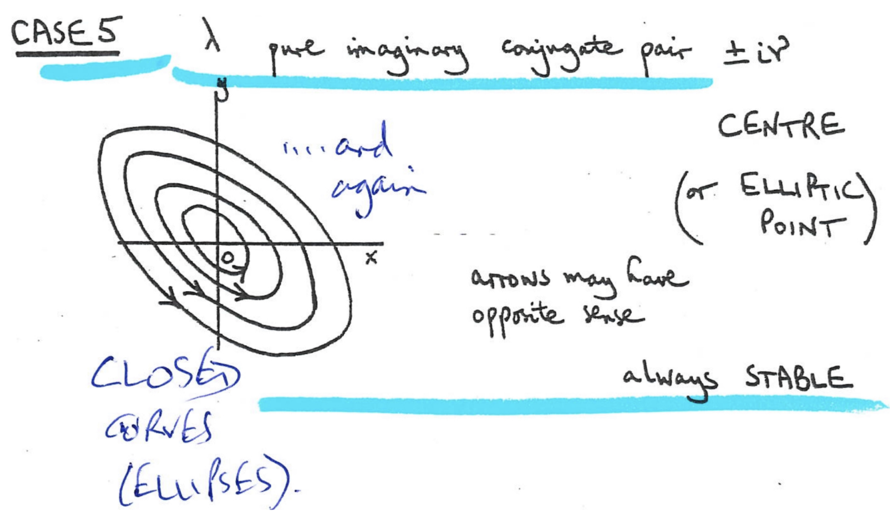


Figure 2.11: case 5

Further explanation for case 5: say $x(t)$ and $y(t)$ can be expressed as

$$\begin{aligned} x(t) &= A_1 \cos \nu t + A_2 \sin \nu t \\ y(t) &= ()A_1 \cos \nu t + ()A_2 \sin \nu t \end{aligned}$$

By solving for $\cos \nu t$ and $\sin \nu t$, we can see that they are directly proportional to the different linear combination of $x(t)$ and $y(t)$. By applying the formula $\cos^2 \nu t + \sin^2 \nu t = 1$, we derive that the equation for x and y are both at power 2, therefore we can see that the graph should be *elliptic*.

Example 95.

(i) (2.17) is 'case 2'.

(ii) (2.16) — the damped harmonic oscillator — is:

- $k = 0$ corresponds to 'case 5'.
- $k^2 < \omega^2$ coresponds to 'case 4'.
- $k^2 > \omega^2$ corresponds to 'case 1'.
- $k^2 = \omega^2$ corresponds to 'case 3(ii)'.

(iii) $\lambda_1 = \lambda_2$ awkward (case 3(ii)).

Try solving

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \tag{2.19}$$

$\Rightarrow \lambda_1 = 2 = \lambda_2$ and we have $\mathbf{V} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ (<u>sole</u> eigenvector here), so that $\begin{pmatrix} x \\ y \end{pmatrix} = B_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{2t}$ is certainly part of our solutions. To find the complete solution we can avoid some rather more advanced linear algebra by just anticipating a general form for our second part of the solution

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} te^{2t} + \begin{pmatrix} c \\ d \end{pmatrix} e^{2t}$$

and find a, b, c, d to fit! Note that $\begin{pmatrix} c \\ d \end{pmatrix}$ is needed since it does not span the space! So $\begin{pmatrix} c \\ d \end{pmatrix}$ is not parallel (or antiparallel) to the eigenvector

$\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Substitute into (2.19):

$$2\begin{pmatrix} a \\ b \end{pmatrix} te^{2t} + \left[\begin{pmatrix} a \\ b \end{pmatrix} + 2\begin{pmatrix} c \\ d \end{pmatrix}\right] e^{2t} = M\left[\begin{pmatrix} a \\ b \end{pmatrix} te^{2t} + \begin{pmatrix} c \\ d \end{pmatrix} e^{2t}\right].$$

Equate coefficients respectively of te^{2t}, e^{2t} on each side:

$$\begin{cases} (M-2I)\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ (M-2I)\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \end{cases} \Rightarrow \begin{cases} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{V} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ or any multiple} \\ \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{cases}$$

$$\Rightarrow \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} K \\ -1-K \end{pmatrix} \text{ (K is arbitrary)}$$

Finally we have for this second solution:

$$B_2\begin{pmatrix} 1 \\ -1 \end{pmatrix} te^{2t} + B_2\left[\begin{pmatrix} 0 \\ -1 \end{pmatrix} + K\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right] e^{2t}.$$

The term with coefficient K is not needed since it has already appeared in the first half of the solution. Thus the general solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = B_1\begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{2t} + B_2\left[\begin{pmatrix} 1 \\ -1 \end{pmatrix} te^{2t} + \begin{pmatrix} 0 \\ -1 \end{pmatrix} e^{2t}\right]$$

with two arbitrary constants as required.

(iv) $\lambda_1 \neq \lambda_2$ but complex pair (case 4/5)

(a)

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & 1 \\ -1 & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

$\Rightarrow \lambda_1 = -\frac{1}{2}+i, \lambda_2 = -\frac{1}{2}-i$, and we can find eigenvectors formally

$$\mathbf{V}_1 = \begin{pmatrix} 1 \\ i \end{pmatrix}, \mathbf{V}_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

So we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = B_1\begin{pmatrix} 1 \\ i \end{pmatrix} e^{(-\frac{1}{2}+i)t} + B_2\begin{pmatrix} 1 \\ -i \end{pmatrix} e^{(-\frac{1}{2}-i)t}$$

where B_1, B_2 are arbitrary. Looking at $\Re$ and $\Im$ parts,

$$\begin{aligned}\begin{pmatrix}1\\ \pm i\end{pmatrix} e^{(-\frac{1}{2}\pm i)t} &= \begin{pmatrix}1\\ \pm i\end{pmatrix} e^{-\frac{1}{2}t}(\cos t \pm i\sin t)\\ &= \begin{pmatrix}e^{-\frac{1}{2}t}\cos t\\ -e^{-\frac{1}{2}t}\sin t\end{pmatrix} \pm i \begin{pmatrix}e^{-\frac{1}{2}t}\sin t\\ e^{-\frac{1}{2}t}\cos t\end{pmatrix}\\ \Rightarrow \begin{pmatrix}x\\ y\end{pmatrix} &= C_1 \begin{pmatrix}e^{-\frac{1}{2}t}\cos t\\ -e^{-\frac{1}{2}t}\sin t\end{pmatrix} + C_2 \begin{pmatrix}e^{-\frac{1}{2}t}\sin t\\ e^{-\frac{1}{2}t}\cos t\end{pmatrix}\end{aligned}$$

where $C_1 = B_1 + B_2, C_2 = (B_1 - B_2)i$. The form above is best for real initial condition.

(b) Simple harmonic oscillator ($k = 0$)

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}0 & 1\\ -\omega^2 & 0\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} \iff \frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \omega^2 x = 0, y = \frac{\mathrm{d}x}{\mathrm{d}t}.$$

$\Rightarrow \lambda_1 = i\omega, \lambda_2 = -i\omega,\ \mathbf{V}_1 = \begin{pmatrix}1\\ i\omega\end{pmatrix}, \mathbf{V}_2 = \begin{pmatrix}1\\ -i\omega\end{pmatrix}$. Therefore,

$$\begin{aligned}\begin{pmatrix}x\\ y\end{pmatrix} &= B_1\begin{pmatrix}1\\ i\omega\end{pmatrix} e^{i\omega t} + B_2 \begin{pmatrix}1\\ -i\omega\end{pmatrix} e^{-i\omega t}\\ &= C_1\begin{pmatrix}\cos\omega t\\ -\omega\sin\omega t\end{pmatrix} + C_2\begin{pmatrix}\sin\omega t\\ \omega\cos\omega t\end{pmatrix}.\end{aligned}$$

2.3.4 Extensions

(i) Inhomogeneous systems

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}x\\ y\end{pmatrix} - M\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}f_1(t)\\ f_2(t)\end{pmatrix}$$

can be solved by the standard method we used for simple linear ordinary differential equations

$$\begin{pmatrix}x\\ y\end{pmatrix}_{\mathrm{GS}} = \begin{pmatrix}x\\ y\end{pmatrix}_{\mathrm{CF}} + \begin{pmatrix}x\\ y\end{pmatrix}_{\mathrm{PI}}$$

We solved CF earlier in this chapter. For PI, it is very difficult in general. Here we just need to find any single particular solution for simple cases such as f_1 and f_2 are constants, then let PI be a constant vector, and solve for it will do.

(ii) Higher order systems

The whole process may be generalised, e.g.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&1&2\\1&2&1\\2&1&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

Since $\det(M-\lambda I)=0$, $\lambda^3-4\lambda^2-\lambda+4=0$,

$$\lambda_1=4,\lambda_2=1,\lambda_3=-1,\mathbf{V}_1=\begin{pmatrix}1\\1\\1\end{pmatrix},\mathbf{V}_2=\begin{pmatrix}1\\-2\\1\end{pmatrix},\mathbf{V}_3=\begin{pmatrix}1\\0\\-1\end{pmatrix}$$

$$\Rightarrow\begin{pmatrix}x\\y\\z\end{pmatrix}=B_1\begin{pmatrix}1\\1\\1\end{pmatrix}e^{4t}+B_2\begin{pmatrix}1\\-2\\1\end{pmatrix}e^{t}+B_3\begin{pmatrix}1\\0\\-1\end{pmatrix}e^{-t}.$$

It is also possible to have equal roots, or complex conjugate pair!

2.4 Partial Differentiation

2.4.1 Introduction

Consider a function $u=u(x,y)$ of 2 independent variables x,y. We can think of u as being the height of a surface above the (x,y) plane. It is often helpful to visualize the surface using *contour lines*

$$u(x,y)=C$$

for different values of c, which is a constant. The countour lines are marked with the value C which $u(x,y)$ would give.

Physically u could represent a geometrical object or temperature or pressure or ... We now look at (spatial) rates of change. Firstly, start at $P(x,y)$ and move a small distance $\delta x=h$ in the x direction to $Q(x+h,y)$ i.e. keeping y fixed.

Definition 96. We define (if the limit exists):

$$\frac{\partial u}{\partial x}=\lim_{h\to 0}\left[\frac{u(x+h,y)-u(x,y)}{h}\right]$$

as the rate of change of u with respect to x at P (keeping y fixed).

Notations: $\frac{\partial u}{\partial x}$, $\left(\frac{\partial u}{\partial x}\right)_y$, $u_x, \ldots$ Be careful with the subscripts!

Similarly, for $P(x,y) \to P(x, y+k)$, we define:

$$\frac{\partial u}{\partial y} = \lim_{k\to 0}\left[\frac{u(x,y+k)-u(x,y)}{k}\right]$$

as the rate of change of u with respect to y at P (keeping x fixed). The notations: $\frac{\partial u}{\partial y}$, $\left(\frac{\partial u}{\partial y}\right)_x$, $u_y, \ldots$

Examples

(i) $u = x^2 \sin y + y^3$

$$\Rightarrow \frac{\partial u}{\partial x} = 2x\sin y, \quad \frac{\partial u}{\partial y} x^2 \cos y + 3y^2.$$

We can, of course, consider higher derivatives:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = u_{xx}, \quad \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right) = u_{xy}$$

Note the order of the partial differentiation in the second expression. For the example above we have

$$\frac{\partial^2 u}{\partial x^2} = 2\sin y, \qquad \frac{\partial^2 u}{\partial y^2} = -x^2 \sin y + 6y,$$
$$\frac{\partial^2 u}{\partial y \partial x} = 2x\cos y, \quad \frac{\partial^2 u}{\partial x \partial y} = 2x \cos y.$$

We note that, in this case, $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}$, which is a general result, requiring only continuity of LHS and RHS — usually the case. E.g. $\frac{x-y}{x+y}$ is not continuous at $(x,y) = (0,0)$.

Similarly, we generally have $u_{xxyyx} = u_{yxyxx} = \ldots$.

(ii) $u(x,y) = a\sin(x - ct)$

$$\frac{\partial u}{\partial x} = a\cos(x-ct), \qquad \frac{\partial^2 u}{\partial x^2} = -a\sin(x-ct),$$
$$\frac{\partial u}{\partial t} = -ac\cos(x-ct), \quad \frac{\partial^2 u}{\partial t^2} = -ac^2\sin(x-ct).$$

We can see that $u(x,t)$ satisfies

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

which is a 2nd order linear P.D.E. It is the one-dimensional wave equation.

In fact, any reasonable function $f(x-ct)$ will satisfy this equation! It represents a wave form moving (with $c>0$ here) to the right. (if $c<0$, then moving to the left)

(iii) $u = \tan^{-1}\frac{y}{x}$. This satisfies

$$\nabla^2 u = (\nabla\cdot\nabla)u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

which is a second order linear P.D.E. It is famously known as the Laplace's equation.

2.4.2 The Total Differential

When we have a function of a single variable $f(x)$, and we make a small change $x \to x+\delta x$ so that $f \to f+\delta f$, then $\delta f \approx \frac{\mathrm{d}f}{\mathrm{d}x}\delta x$. In the limit ("small, $\to 0$")

$$\mathrm{d}f = \frac{\mathrm{d}f}{\mathrm{d}x}\mathrm{d}x.$$

Now for a function of two variables, small changes $x \to x+\delta x$, $y \to y+\delta y$ lead to $u(x,y) \to u+\delta u$, with $\delta u = u(x+\delta x, y+\delta y) - u(x,y)$, so $\delta u \approx \frac{\partial u}{\partial x}\delta x + \frac{\partial u}{\partial y}\delta y$.

Definition 97. The ***total differential*** of $u(x,y)$ is

$$\mathrm{d}u = \frac{\partial u}{\partial x}\mathrm{d}x + \frac{\partial u}{\partial y}\mathrm{d}y.$$

The idea can be generalized to higher dimensions, e.g. $u(x,y,z)$ etc.

Examples

(i) $u = x^2 \sin y + y^3$, and we can get

$$\delta u \approx (2x \sin y)\delta x + (x^2 \cos y + 3y^2)\delta y$$

$$\mathrm{d}u = (2x \sin y)\mathrm{d}x + (x^2 \cos y + 3y^2)\mathrm{d}y.$$

(ii) Area of a rectangle $A = xy$. Here

$$\begin{aligned}\delta A &= (x + \delta x)(y + \delta y) - xy \\ &= \underbrace{y\delta x + x\delta y}_{\text{1st order small}} + \underbrace{\delta x \delta y}_{\text{2nd order small}}\end{aligned}$$

Therefore $\delta A \approx y\delta x + x\delta y \iff \mathrm{A} = y\mathrm{d}x + x\mathrm{d}y$.

(iii) Height of a building $h = x \tan\theta$.

Let $x = 200$m with error ± 2m, $\theta = 20^\circ$ with error $\pm\frac{1}{2}^\circ$. We can derive that

$$\delta h \approx (\tan\theta)\delta x + (x \sec^2\theta)\delta\theta.$$

Central estimate is $200 \tan\left(\frac{\pi}{9}\right) = 72.8$m.

$$\Rightarrow \delta h \approx 0.36\delta x + 226.5\delta\theta$$

with $|\delta x| \leq 2, |\delta\theta| \leq \frac{\pi}{360} = 0.0087$. So

$$|\delta h| \leq (0.36)(2) + (226.5)(0.0087) = 2.7\text{m}$$

and $h = 72.8 \pm 2.7$m. (which is $\pm 3.7\%$)

2.4.3 Function of a function — ‘The Chain Rule’

If we have $u = f(x)$ and $x = g(t)$, then we have as a consequence

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}u}{\mathrm{d}x} \cdot \frac{\mathrm{d}x}{\mathrm{d}t} = f'(x)g'(t) = f'(g(t))g'(t).$$

Now consider $u = u(x, y)$ where $x(t)$ and $y(t)$. We saw previously that

$$\delta u \approx \frac{\partial u}{\partial x}\delta x + \frac{\partial u}{\partial y}\delta y \quad \Longrightarrow \quad \frac{\delta u}{\delta t} \approx \frac{\partial u}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial u}{\partial y}\frac{\delta y}{\delta t}.$$

Definition 98. The chain rule in 3-dimension is

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\partial u}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial u}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}$$

where x and y are functions *only* of t.

E.g. if $x(r, s)$ and $y(r, s)$ then

$$\frac{\partial \overline{u}}{\partial r} = \frac{\partial \overline{u}}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial \overline{u}}{\partial y}\frac{\partial y}{\partial r}, \qquad \frac{\partial \overline{u}}{\partial s} = \frac{\partial \overline{u}}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial \overline{u}}{\partial y}\frac{\partial y}{\partial s}.$$

Examples

(i) Volume V of a cylindrical box of radius r and height h: $V = \pi r^2 h$.

If we know that $r = 2t, h = 1 + t^2$, then

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}t} &= \frac{\partial V}{\partial r}\frac{\mathrm{d}r}{\mathrm{d}t} + \frac{\partial V}{\partial h}\frac{\mathrm{d}h}{\mathrm{d}t}\\ &= (2\pi r h)(2) + (\pi r^2)(2t)\\ &= 8\pi(t + 2t^3)\end{aligned}$$

<u>Check</u>: $V = \pi(2t)^2(1+t^2) = 4\pi(t^2+t^4)$, and of course $\frac{\mathrm{d}V}{\mathrm{d}t} = 8\pi t + 16\pi t^3$.

(ii) $u(x, y) = \overline{u}(s, t) = x^2 y$ with $x = st$, $y = s + t$.

$$\Rightarrow \begin{cases} \dfrac{\partial \overline{u}}{\partial s} = \dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial s} + \dfrac{\partial u}{\partial y}\dfrac{\partial y}{\partial s} = (2xy)(t) + (x^2)(1) = \cdots = 3s^2t^2 + 2st^3 \\ \dfrac{\partial \overline{u}}{\partial t} = \dfrac{\partial u}{\partial x}\dfrac{\partial x}{\partial t} + \dfrac{\partial u}{\partial y}\dfrac{\partial y}{\partial t} = (2xy)(s) + (x^2)(1) = \cdots = 3s^2t^2 + 2s^3t. \end{cases}$$

Again, we can check this since $u(x, y) = (st)^2(s + t) = \overline{u}(s, t)$.

(iii) $u(x, y) = xy + y^2$, a *cautionary* example!

If we have a second relation $y = x + t$, then of course

$$\overline{u}(x, t) = x(x + t) + (x + t)^2.$$

We have to be careful when we look at the variation of u and $\overline{u}$ with respect to x:

(a) $\frac{\partial u}{\partial x} = y(= x + t)$
(b) $\frac{\partial \overline{u}}{\partial x} = 2x + t + 2(x + t)(= 4x + 3t)$.

These are not the same — u and $\overline{u}$ have the same function values at corresponding points. But in (a), y is being kept constant $\left(\frac{\partial u}{\partial x}\right)_y$, and in (b), t is being kept constant $\left(\frac{\partial \overline{u}}{\partial x}\right)_t$, So care is needed evidently! Ensure that the substitution is *disjoint*!

2.4.4 From Cartesians to Polars

$$\underbrace{u(x,y)}_{\text{Cartesians}} = \underbrace{\overline{u}(r,\theta)}_{\text{Plane Polars}}$$

with

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}, \qquad \begin{cases} r = (x^2 + y^2)^{\frac{1}{2}} \\ \theta = \tan^{-1}\dfrac{y}{x} \end{cases}.$$

Both are *orthogonal* systems! We need to be careful! E.g.

$$\left(\frac{\partial x}{\partial r}\right)_\theta = \cos\theta \quad \text{and} \quad \left(\frac{\partial r}{\partial x}\right)_y = \frac{x}{(x^2 + y^2)^{\frac{1}{2}}} = \cos\theta.$$

They are keeping different constants, so we should *not* be tempted!

$$\left(\frac{\partial x}{\partial r}\right)_\theta \neq \frac{1}{\left(\frac{\partial r}{\partial x}\right)_y}.$$

<u>Note</u>: The Cartesian and Polar versions of our function have the same function values, but are described differently! E.g.

$$u(x,y) = x^2 + y^2 = r^2 = \overline{u}(r,\theta)$$

and it is not $(r^2 + \theta^2)$.

Chain rule

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial \overline{u}}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial \overline{u}}{\partial \theta}\frac{\partial \theta}{\partial x} \\ &= (\cos\theta)\frac{\partial \overline{u}}{\partial r} + \left(-\frac{\sin\theta}{r}\right)\frac{\partial \overline{u}}{\partial \theta} \end{aligned}$$

and

$$\begin{aligned}\frac{\partial u}{\partial y} &= \frac{\partial \overline{u}}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial \overline{u}}{\partial \theta}\frac{\partial \theta}{\partial y} \\ &= (\sin\theta)\frac{\partial \overline{u}}{\partial r} + \left(\frac{\cos\theta}{r}\right)\frac{\partial \overline{u}}{\partial \theta}.\end{aligned}$$

Definition 99. The ***partial differential operators*** are defined as

$$\begin{aligned}\frac{\partial}{\partial x} &= \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta} \\ \frac{\partial}{\partial y} &= \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}\end{aligned}$$

which relate rates of change in the two different coordinate systems!

Examples

(i) $u(x, y) = x^2 - y^2 = r^2(\cos^2\theta - \sin^2\theta) = \overline{u}(r, \theta)$.

$$\frac{\partial u}{\partial x} = 2x = 2r\cos\theta, \quad \frac{\partial u}{\partial y} = -2y = -2r\sin\theta.$$

(ii) We can express $\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2$ in plane polars as

$$\begin{aligned}\left[(\cos\theta)\frac{\partial \overline{u}}{\partial r} + \left(-\frac{\sin\theta}{r}\right)\frac{\partial \overline{u}}{\partial \theta}\right]^2 + \left[(\sin\theta)\frac{\partial \overline{u}}{\partial r} + \left(\frac{\cos\theta}{r}\right)\frac{\partial \overline{u}}{\partial \theta}\right]^2 \\ = \left(\frac{\partial \overline{u}}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial \overline{u}}{\partial \theta}\right)^2.\end{aligned}$$

(iii) For Laplace equation $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y} = 0$,

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \left(\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\right)\left(\cos\theta\frac{\partial \overline{u}}{\partial r} - \frac{\sin\theta}{r}\frac{\partial \overline{u}}{\partial \theta}\right) \\ &\vdots \quad \text{Attrition!} \\ &= \cos^2\theta\frac{\partial^2 \overline{u}}{\partial r^2} - \frac{2\cos\theta\sin\theta}{r}\frac{\partial^2 \overline{u}}{\partial r \partial \theta} + \frac{\sin^2\theta}{r}\frac{\partial \overline{u}}{\partial r} \\ &\qquad + \frac{\sin^2\theta}{r^2}\frac{\partial^2 \overline{u}}{\partial \theta^2} + \frac{2\sin\theta\cos\theta}{r^2}\frac{\partial \overline{u}}{\partial \theta}\end{aligned}$$

and

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right) = \left(\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial\theta}\right)\left(\sin\theta\frac{\partial\overline{u}}{\partial r} + \frac{\cos\theta}{r}\frac{\partial\overline{u}}{\partial\theta}\right)$$

$\vdots$ Attrition!

$$= \sin^2\theta\frac{\partial^2\overline{u}}{\partial r^2} + \frac{2\cos\theta\sin\theta}{r}\frac{\partial^2\overline{u}}{\partial r\partial\theta} + \frac{\cos^2\theta}{r}\frac{\partial\overline{u}}{\partial r} + \frac{\cos^2\theta}{r^2}\frac{\partial^2\overline{u}}{\partial\theta^2} - \frac{2\sin\theta\cos\theta}{r^2}\frac{\partial\overline{u}}{\partial\theta}.$$

Hence

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \underbrace{\frac{\partial^2\overline{u}}{\partial r^2} + \frac{1}{r}\frac{\partial\overline{u}}{\partial r}}_{=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\overline{u}}{\partial r}\right)} + \frac{1}{r^2}\frac{\partial^2\overline{u}}{\partial\theta^2} = 0.$$

Laplace in 3 dimensions is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0.$$

2.4.5 Implicit Functions

If we have a function defined implicitly $F(x, y) = 0$, then F does not change as x and y do so. The total derivative then is

$$\mathrm{d}F = \frac{\partial F}{\partial x}\mathrm{d}x + \frac{\partial F}{\partial y}\mathrm{d}y = 0.$$

So the derivative of y with respect to x is given by

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{F_x}{F_y}.$$

Examples

(i) $F(x, y) = x^2\sin y + xy - 1 = 0$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{2x\sin y + y}{x^2\cos y + x}.$$

If we have an implicit function of 3 variables $F(x, y, z) = 0$, this constrains our point (x, y, z) to be on a particular surface. We can certainly regard, if we wish, $x = x(y, z)$ or $y = y(x, z)$ or $z = z(x, y)$. Now, no change in F on the surface,

$$\Rightarrow \mathrm{d}F = \frac{\partial F}{\partial x}\mathrm{d}x + \frac{\partial F}{\partial y}\mathrm{d}y + \frac{\partial F}{\partial z}\mathrm{d}z = 0.$$

Then:

- At constant y: $\left(\frac{\partial z}{\partial x}\right)_y = -\frac{F_x}{F_z}$
- At constant x: $\left(\frac{\partial z}{\partial y}\right)_x = -\frac{F_y}{F_z}$
- At constant z: $\left(\frac{\partial y}{\partial x}\right)_z = -\frac{F_x}{F_y}$.

<u>Note</u>: Here

$$\left(\frac{\partial z}{\partial x}\right)_y = \frac{1}{\left(\frac{\partial x}{\partial z}\right)_y}$$

because the variable y is being kept constant on both sides — we are looking at variation on a constant y slice of the $F = 0$ surface!

(ii) In thermodynamics the equation of *state* of a gas/liquid is written

$$F(p, V, T) = 0$$

It is an implicit definition of $p = p(V, T)$. Only in simple cases can we express this relation *explicitly*, e.g. ideal gas $P = \frac{RT}{V}$.

In any case, from the general relation above, we can show that

$$\left(\frac{\partial P}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial T}{\partial P}\right)_V = -1$$

an example of an *exact thermodynamic identity*.

2.4.6 Taylor Series

Recall that for functions with *one* independent variable, we have

$$u(x) = u(x_0) + (x - x_0)u'(x_0) + \frac{(x - x_0)^2}{2!}u''(x_0) + \cdots$$

or

$$u(x_0+h) = u(x_0) + hu'(x_0) + \frac{h^2}{2!}u''(x_0) + \cdots$$

and it is called 'Maclaurin' if $x_0 = 0$.

Now extended: we now consider a function $u(x, y)$ of two independent variables x, y in the neighbourhood of (x_0, y_0). That is we seek an expansion in power of $h = x - x_0$, $k = y - y_0$. So

$$\begin{aligned} u(x,y) &= u(x_0+h, y_0+k) \\ &= u(x_0, y_0+k) + h\frac{\partial}{\partial x}[u(x_0,y_0+k)] + \frac{h^2}{2!}\frac{\partial^2}{\partial x^2}[u(x_0,y_0+k)] + \cdots \end{aligned}$$

where

$$\begin{aligned} u(x_0,y_0+k) &= u(x_0,y_0) + k\frac{\partial}{\partial y}u(x_0,y_0) + \frac{k^2}{2!}\frac{\partial^2}{\partial y^2}[u(x_0,y_0)] + \cdots \\ h\frac{\partial}{\partial x}[u(x_0,y_0+k)] &= h\left[\frac{\partial}{\partial x}[u(x_0,y_0)] + k\frac{\partial^2}{\partial y \partial x}[u(x_0,y_0)] + \cdots\right] \\ \frac{h^2}{2!}\frac{\partial^2}{\partial x^2}[u(x_0,y_0+k)] &= \frac{h^2}{2!}\left[\frac{\partial^2}{\partial x^2}[u(x_0,y_0)] + k\frac{\partial^3}{\partial y \partial x^2}[u(x_0,y_0)] + \cdots\right] \\ &\vdots \end{aligned}$$

Now collect the terms, and we get

$$\begin{aligned} u(x,y) = u_0 &+ \underbrace{\left[h\left(\frac{\partial u}{\partial x}\right)_0 + k\left(\frac{\partial u}{\partial y}\right)_0\right]}_{\text{first order}} \\ &+ \frac{1}{2!}\underbrace{\left[h^2\left(\frac{\partial^2 u}{\partial x^2}\right)_0 + 2hk\left(\frac{\partial^2 u}{\partial x \partial y}\right)_0 + k^2\left(\frac{\partial^2 u}{\partial y^2}\right)_0\right]}_{\text{second order}} + \cdots \end{aligned} \tag{2.20}$$

where the subscript 0 means that it is evaluated at (x_0, y_0).

There is a straightforward pattern to these terms, where we are assuming that $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}$ etc. We can write the operator

$$\mathscr{D} = h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}.$$

The above Taylor series can then be re-written as

$$u(x_0+h, y_0+k) = u_0 + \mathscr{D}u_0 + \frac{\mathscr{D}^2 u_0}{2!} + \frac{\mathscr{D}^3 u_0}{3!} + \cdots$$

and our pattern involves the binomial coefficients explicitly, i.e.

$$\mathscr{D}^2 = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right) = h^2\frac{\partial^2}{\partial x^2} + 2hk\frac{\partial^2}{\partial x \partial y} + k^2\frac{\partial^2}{\partial x^2},$$

$$\mathscr{D}^3 u_0 = h^3\frac{\partial^3 u_0}{\partial x^3} + 3h^2 k\frac{\partial^3 u_0}{\partial x^2 \partial y} + 3hk^2\frac{\partial^3 u_0}{\partial x \partial y^2} + k^3\frac{\partial^3 u_0}{\partial y^3}, \text{ etc.}$$

Example

$$u(x,y) = e^{2x-y}$$

where $x_0 = 0 = y_0, x = 0+h, y = 0+k, u_0 = u(0,0) = 1$. So

$$\frac{\partial u}{\partial x} = 2e^{2x-y}, \frac{\partial u}{\partial y} = -e^{2x-y} \Rightarrow \left(\frac{\partial u}{\partial x}\right)_0 = 2, \left(\frac{\partial u}{\partial y}\right)_0 = -1.$$

$$\frac{\partial^2 u}{\partial x^2} = 4e^{2x-y}, \frac{\partial^2 u}{\partial x \partial y} = -2e^{2x-y}, \frac{\partial^2 u}{\partial y^2} = e^{2x-y},$$

$$\Rightarrow \left(\frac{\partial^2 u}{\partial x^2}\right)_0 = 4, \left(\frac{\partial^2 u}{\partial x \partial y}\right)_0 = -2, \left(\frac{\partial^2 u}{\partial y^2}\right)_0 = 1.$$

Thus

$$e^{2x-y} = e^{2h-k} = 1 + (2h-k) + \frac{1}{2!}[4h^2 - 4hk + k^2] + \cdots.$$

Check: $e^{2h-k} = 1 + (2h-k) + \frac{1}{2!}(2h-k)^2 + \cdots$.

2.4.7 Stationary Points

We plotted surfaces $u(x,y)$ and their contours, we now ask whether our surface has a *horizontal tangent plan* at any point.

For *one* independent variable, we have local maximum, local minimum, and inflexion point with horizontal tangent as the three cases for having horizontal tangent line. For *two* independent variable, we also have three types of stationary points each with a local horizontal tangent plane.

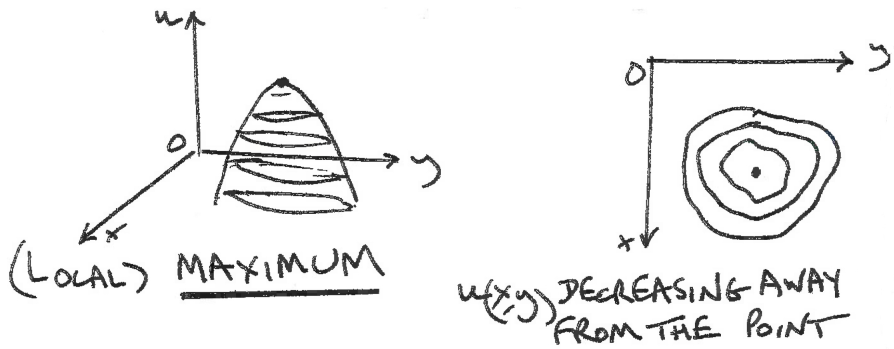


Figure 2.12: local maximum

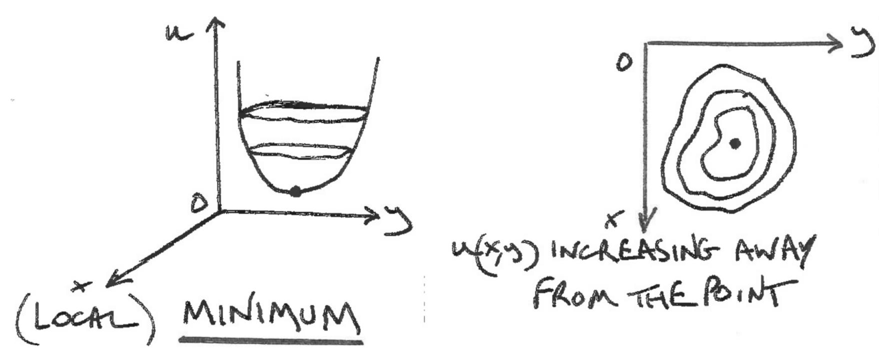


Figure 2.13: local minimum

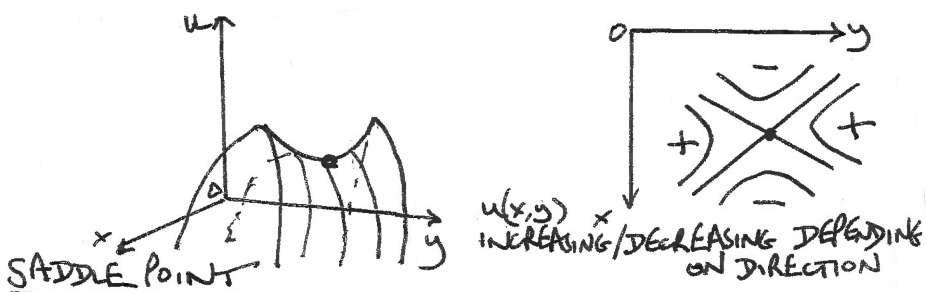


Figure 2.14: saddle point

How do we distinguish between these cases?

Consider a stationary point (x_0, y_0) where we have (of course) $\left(\frac{\partial u}{\partial x}\right)_0 = 0 = \left(\frac{\partial u}{\partial y}\right)_0$. Then we write out the Taylor expansion for $u(x, y)$ about (x_0, y_0) exactly as Equation (2.20). The first order terms are cancelled out, so

$$\begin{aligned} \delta u &= u(x_0 + h, y_0 + k) - u(x_0, y_0) \\ &= \frac{1}{2}\left[Ah^2 + 2Bhk + Ck^2\right] + \cdots \end{aligned}$$

where

$$A = \left(\frac{\partial^2 u}{\partial x^2}\right)_0, \quad B = \left(\frac{\partial^2 u}{\partial x \partial y}\right)_0, \quad C = \left(\frac{\partial^2 u}{\partial y^2}\right)_0,$$

and δu can be written in matrix form:

$$\delta u = \frac{1}{2}\begin{pmatrix} h & k \end{pmatrix}\begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix} + \cdots .$$

<u>Note</u>: If A, B, C are all 0, then we need higher-order terms.

As such,

(i) $\delta u > 0$ for *any* small h, k then $u(x_0, y_0)$ is a (local) minimum.

(ii) $\delta u < 0$ for *any* small h, k then $u(x_0, y_0)$ is a (local) maximum.

(iii) δu can be positive or negative depending on the value of h, k, then $u(x_0, y_0)$ is a saddle point.

The easiest way to determine this is via e.g.

$$\delta u = \frac{1}{2}k^2\left[A\left(\frac{h}{k}\right)^2 + 2B\left(\frac{h}{k}\right) + C\right] + \cdots$$

and consider

$$F(\lambda) = A\lambda^2 + 2B\lambda + C.$$

(Possible to think in terms of $B^2 - AC$, as this is more "natural" while thinking about *discriminant* of polynomials. Then all the following deductions/conclusions should be reversed!)

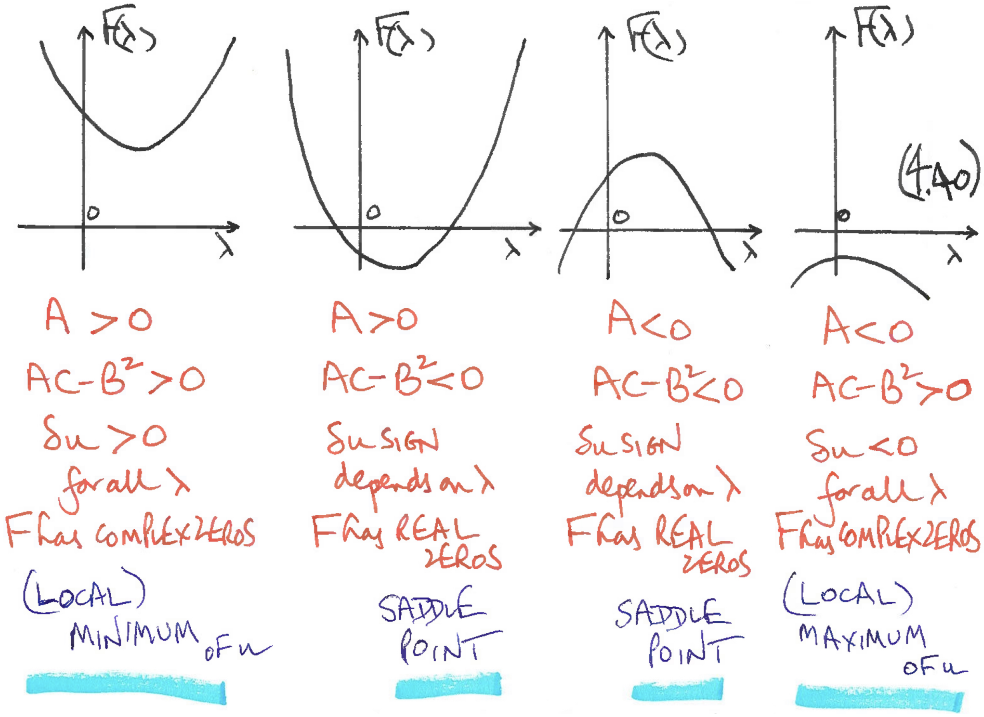


Figure 2.15: graphs to determine what type of stationary point it is

Summary For functions of two variables $u(x, y)$, stationary points are located at simultaneous solution of the two equations:

$$\begin{cases} \dfrac{\partial u}{\partial x} = 0 \\ \dfrac{\partial u}{\partial y} = 0 \end{cases}$$

where $\mathrm{d}u = 0$ locally. Each (x_0, y_0) has *character* determined by

$$E_0 = \left[\left(\frac{\partial^2 u}{\partial x^2}\right)\left(\frac{\partial^2 u}{\partial y^2}\right) - \left(\frac{\partial^2 u}{\partial x \partial y}\right)^2 \right]_0 = AC - B^2$$

with

- $E_0 < 0$ implies a saddle point,
- $E_0 > 0$ and
  - $\left(\frac{\partial^2 u}{\partial x^2}\right)_0 < 0$ implies a (local) maximum,
  - $\left(\frac{\partial^2 u}{\partial x^2}\right)_0 > 0$ implies a (local) minimum.

Of course there are *singular* cases that can occur, such as $E_0 = 0, A = B = C = 0$ etc. These normally require higher-order derivatives to determine the issue — not considered here.

Examples

(i) $u(x, y) = x^3 + xy^2 - x - yx^2 - y^3 + y = (x - y)(x^2 + y^2 - 1)$.

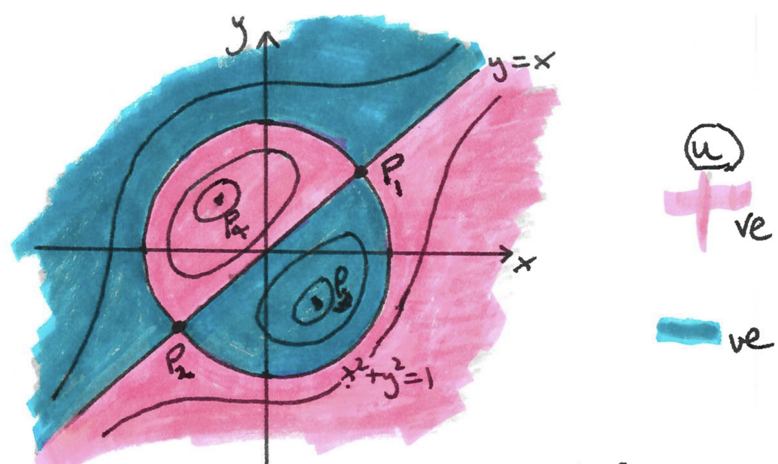
$$\Rightarrow \begin{cases} \dfrac{\partial u}{\partial x} = 3x^2 + y^2 - 1 - 2xy = 0 \\ \dfrac{\partial u}{\partial y} = 2xy - x^2 - 3y^2 + 1 = 0. \end{cases}$$

By adding the two equations up, we get $2x^2 - 2y^2 = 0, y = \pm x$. When $y = x$, $x = \pm\frac{1}{\sqrt{2}}$; when $y = -x$, $x = \pm\frac{1}{\sqrt{6}}$. Thus there are four stationary points:

$$P_1\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), P_2\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), P_3\left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right), P_4\left(-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right).$$

Pt	$A=\left(\frac{\partial^2 u}{\partial x^2}\right)_0 = (6x-2y)_0$	$B=\left(\frac{\partial^2 u}{\partial x \partial y}\right)_0 = (2y-2x)_0$	$C=\left(\frac{\partial^2 u}{\partial y^2}\right)_0 = (2x-6y)_0$	$AC-B^2 \equiv E_0$	u_0	TYPE
P_1	$4/\sqrt{2}$	0	$-4/\sqrt{2}$	-8	0	SADDLE
P_2	$-4/\sqrt{2}$	0	$4/\sqrt{2}$	-8	0	SADDLE
P_3	$8/\sqrt{6}$	$-4/\sqrt{6}$	$8/\sqrt{6}$	$+8$	$-\frac{2}{3}\sqrt{\frac{2}{3}}$	MINIMUM
P_4	$-8/\sqrt{6}$	$4/\sqrt{6}$	$-8/\sqrt{6}$	$+8$	$+\frac{2}{3}\sqrt{\frac{2}{3}}$	MAXIMUM

Figure 2.16: stationary points analysis



The ZERO CONTOUR is $y=x$ together with $x^2+y^2=1$.

Figure 2.17: contour sketch

Then we can analyse the stationary points, as shown in Figure 2.16, and sketch the contours, as shown in Figure 2.17!

Warning: When we are faced with a function of several variables and we need to find stationary points (and potential local max and min), we need to ensure that our *independent* variables are indeed independent.

(ii) Maximise volume $V = xyz$ of a rectangular box given the surface area $A = 2xy + 2yz + 2zx$ is fixed. In this case, x, y, z are *not* independent. We then need to write

$$z = \frac{A - 2xy}{2(x+y)} \Rightarrow V = \frac{xy(A - 2xy)}{2(x+y)}.$$

Now x, y are independent, so we can solve and derive that

$$x_0 = \sqrt{\frac{A}{6}} = y_0 (= z_0), V_{\max} = \left(\frac{A}{6}\right)^{\frac{3}{2}}.$$

2.4.8 Application — Exact (First Order) Differential Equations

We know that for a function of two variables $u(x, y)$ the total differential is

$$\mathrm{d}u = \frac{\partial u}{\partial x}\mathrm{d}x + \frac{\partial u}{\partial y}\mathrm{d}y$$

and of course $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ will both, in general, be functions of x and y. Now consider the converse problem! Given

$$P(x, y)\mathrm{d}x + Q(x, y)\mathrm{d}y$$

i.e. given P and Q, when is it the case that this *is* the total differential of some (as yet unknown) function $u(x, y)$? If it is such then $P(x, y) = \frac{\partial u}{\partial x}$ and $Q(x, y) = \frac{\partial u}{\partial y}$ for that function $u(x, y)$. This implies (easy!) and is implied by (proof not given here!) the *condition of integrability*:

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

Examples

(i) $y^2 \mathrm{d}x + (x^2 + 2y)\mathrm{d}y$. Since

$$\frac{\partial P}{\partial y} = 2y \not\equiv \frac{\partial Q}{\partial x} = 2x,$$

the test fails, and thus not an exact/total differential.

(ii) $(2xy + \cos x \cos y)\mathrm{d}x + (x^2 - \sin x \sin y)\mathrm{d}y$.

$$\frac{\partial P}{\partial y} = 2x - \cos x \sin y = \frac{\partial Q}{\partial x},$$

so the test pass and it is exact. Thus

$$\frac{\partial u}{\partial x} = 2xy + \cos x \cos y$$

$$\Rightarrow u(x, y) = x^2 y + \sin x \cos y + f(y)$$

where $f(y)$ is a 'constant' of integration w.r.t. x. Then either

$$\begin{aligned} \frac{\partial u}{\partial y} &= x^2 - \sin x \sin y + \frac{\mathrm{d}f(y)}{\mathrm{d}y} \\ &= x^2 - \sin x \sin y \end{aligned}$$

so that $\frac{\mathrm{d}f}{\mathrm{d}y} = 0$ and $f(y) = K$ constant w.r.t. x and y, or alternatively

$$\frac{\partial u}{\partial y} = x^2 - \sin x \sin y$$

$$\Rightarrow u(x, y) = x^2 y + \sin x \cos y + g(x)$$

where $g(x)$ is the 'constant' of integration w.r.t. y. Comparing the two expressions and deduce that $f(y) = g(x) = K$ constant independent of x and y. This also gives

$$u(x, y) = x^2 y + \sin x \cos y + K$$

and $P\mathrm{d}x + Q\mathrm{d}y = \mathrm{d}u(x, y)$.

Now consider an equation of the form

$$P(x,y)\mathrm{d}x + Q(x,y)\mathrm{d}y = 0. \tag{2.21}$$

This is a first order differential equation, with alternative forms $\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{P(x,y)}{Q(x,y)}$ and $P + Q\frac{\mathrm{d}y}{\mathrm{d}x} = 0$. If $P\mathrm{d}x + Q\mathrm{d}y$ is the total differential of a function $u(x,y)$, i.e. $\mathrm{d}u(x,y) = 0$, then the solution is

$$u(x,y) = \text{ constant } C.$$

In this case (2.21) is an ***exact differential equation***. From the example (ii) above,

$$(2xy + \cos x \cos y)\mathrm{d}x + (x^2 - \sin x \sin y)\mathrm{d}y = 0$$

has solution

$$u(x,y) = x^2 y + \sin x \cos y = C.$$

What if $P\mathrm{d}x + Q\mathrm{d}y$ is not exact, which would mean that (2.21) would not be an exact differential equation? We can consider multiplying through by a factor $\lambda(x,y)$,

$$\Rightarrow (\lambda P)\mathrm{d}x + (\lambda Q)\mathrm{d}y = 0$$

which is the 'same' differntial equation as (2.21), but can we make it exact? Evidently we would need

$$\frac{\partial}{\partial y}(\lambda P) = \frac{\partial}{\partial x}(\lambda Q) \tag{2.22}$$

$$\Rightarrow P\frac{\partial \lambda}{\partial y} - Q\frac{\partial \lambda}{\partial x} + \lambda\left(\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x}\right) = 0.$$

This is a *partial* differential equation for λ given P, Q. It can be shown that there is a solution, but this isn't normally very helpful in finding it! However, sometimes we can find a suitable λ by *inspection*!

Example

(1)

$$(xy - 1)\mathrm{d}x + (x^2 - xy)\mathrm{d}y = 0$$

$$\Rightarrow \frac{\partial P}{\partial y} = x \not\equiv \frac{\partial Q}{\partial x} = 2x - y.$$

Let's *try* $\lambda(x)$ (with x only) say and use (2.22),

$$\Rightarrow \lambda(x)x = \frac{\mathrm{d}\lambda(x)}{\mathrm{d}x}(x^2 - xy) + \lambda(x)(2x - y)$$

($\frac{\mathrm{d}\lambda(x)}{\mathrm{d}y} = 0$) so that $\frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}x} = -\frac{1}{x}$, $\lambda = \frac{K}{x}$, and we can take $K = 1$ W.L.O.G.

$$\Rightarrow (y - \frac{1}{x})\mathrm{d}x + (x - y)\mathrm{d}y = 0 \quad \text{is exact!}$$

$$\Rightarrow u(x, y) = xy - \ln|x| - \frac{1}{2}y^2 - C,$$

$$xy - \ln|x| - \frac{1}{2}y^2 = C$$

is the general solution of our equation.

(2)

$$\frac{\mathrm{d}y}{\mathrm{d}x} + f(x)y = g(x)$$

$$\Rightarrow [yf(x) - g(x)]\mathrm{d}x + \mathrm{d}y = 0$$

and $\frac{\partial}{\partial y}(yF - G) = F \not\equiv \frac{\partial}{\partial x}(1) = 0$ in general. Try integrating factor $\lambda(x)$ only:

$$\frac{1}{\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}x} = f(x) \Rightarrow \lambda = K\exp\left[\int^x f(x)\mathrm{d}x\right].$$

So the term *integrating factor* has the same meaning, as in the previous chapter!

2.4.9 Application — Vector Calculus

We often need to consider *scalar* and *vector* functions of position $\mathbf{r}(= x\mathbf{i} + y\mathbf{j} + z\mathbf{k})$ (*fields*): $\phi(\mathbf{r})$ and $\mathbf{u}(\mathbf{r}) = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$. These could represent e.g. the temperature (ϕ) and velocity ($\mathbf{u}$) within a material. Of course they might also practically depend on time t too, but for now we consider *spatial* rates of change.

Definition 100. The ***partial differential operator*** is defined as

$$\nabla \equiv \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$$

There are 3 important derived fields, each involving the (partial) differential operator. They are:

(1)
$$\nabla\phi = \mathbf{i}\frac{\partial\phi}{\partial x} + \mathbf{j}\frac{\partial\phi}{\partial y} + \mathbf{k}\frac{\partial\phi}{\partial z} \equiv \text{ grad } \phi \quad \text{‘Gradient’}$$

which is a vector field. It is measuring *the rate of change of a value in each dimension*.

(2)
$$\nabla\cdot\mathbf{u} = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial u_3}{\partial z} \equiv \text{ div } \mathbf{u} \quad \text{‘Divergence’}$$

which is a scalar field. It is measuring *the overall rate of change of a vector in each standard basis* by decomposing the vector into its corresponding dimensions and summing up the rate of change of each component.

(3)
$$\nabla\times\mathbf{u} = \mathbf{i}\left(\frac{\partial u_3}{\partial y} - \frac{\partial u_2}{\partial z}\right) + \mathbf{j}\left(\frac{\partial u_1}{\partial z} - \frac{\partial u_3}{\partial x}\right) + \mathbf{k}\left(\frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y}\right) \equiv \text{ curl } \mathbf{u}$$

which is a vector field. It is measuring *the rate of rotation/curl about a point*, hence the output is a vector field with vectors describing the direction and magnitude of rotation about each point. The expression can be ‘mnemonically’ memorized in the following form:

$$\det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u_1 & u_2 & u_3 \end{pmatrix}$$

Each of these fields has a strong physical meaning and can be interpreted in e.g. 2 dimensions (x, y) if required.

(1) $\nabla\phi \equiv \text{grad } \phi$.

At each point, $\nabla\phi$ is *perpendicular* to the “ϕ = constant” (say ϕ_1) surface through the point. The direction also increases the value *most rapidly*.

If $\mathrm{d}\mathbf{r} = \mathrm{d}x\mathbf{i} + \mathrm{d}y\mathbf{j} + \mathrm{d}z\mathbf{k}$ along the surface at a point P, then

$$(\nabla\phi) \cdot \mathrm{d}\mathbf{r} = \frac{\partial\phi}{\partial x}\mathrm{d}x + \frac{\partial\phi}{\partial y}\mathrm{d}y + \frac{\partial\phi}{\partial z}\mathrm{d}z = \mathrm{d}\phi = 0$$

because $\phi = \phi_1$ on that surface through P. We note that $\nabla\phi$ is directed towards increasing ϕ.

Example:

(a) $\phi = x^2y + 2xz$.

$$\begin{aligned}\nabla\phi &= (2xy + 2z, x^2, 2x)\\ &= (-2, 4, 4)\end{aligned}$$

at P where $P(2,-2,3)$. Unit normal to surface at $P = \pm\frac{-2\mathbf{i}+4\mathbf{j}+4\mathbf{k}}{\sqrt{36}}$.
The rate of change of ϕ in a given direction $\hat{\mathbf{a}}$ is given by

$$\nabla\phi \cdot \hat{\mathbf{a}} \equiv |\nabla\phi| \cos\alpha$$

where α is the angle between $\nabla\phi$ and $\hat{\mathbf{a}}$.
This is defined as ***directional derivative***: Given a direction and a point, find out the rate of change at that point along the given direction.
This shows that when $\alpha = n\pi$, i.e. the direction perpendicular to the surface, *has the greatest rate of change*, i.e. steepest slope. Following the direction of $2k\pi$ results in the *greatest rate of increment*.
We can use $\nabla\phi$ to find rates of change, normals to curves and surfaces, tangent planes, ...

(b) $2xz^2 - 3xy - 4x - 7 = 0$ (an equation about a surface).
The normal to the surface at $(1,-1,2)$ is

$$(2z^2 - 3y - 4, -3x, 4xz) = 7\mathbf{i} - 3\mathbf{j} + 8\mathbf{k}.$$

Equation of the tangent plane is

$$(\mathbf{r} - \mathbf{r}_0) \cdot (7,-3,8) = 0$$

i.e.

$$\begin{aligned}&((x,y,z) - (1,-1,2)) \cdot (7,-3,8) = 0\\ \Rightarrow\ &7(x-1) - 3(y+1) + 8(z-2) = 0.\end{aligned}$$

(2) $\nabla \cdot \mathbf{u} \equiv \text{div } \mathbf{u}$.

It acts as a measurement of whether a local field is a *source* (+) 'outflow' / *sink* (−) 'inflow'.

From the expression, we can see that the expression adds the rate of change of respective componenets together. It makes sense, and you will feel it so much by listening to how 3B1B explains the intuition, using the following link:

https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/divergence-grant-videos/v/divergence-formula-part-1

In general, the divergence at a point $\mathbf{x}$ is the limit of the ratio of the flux through the surface over the volume enclosing $\mathbf{x}$ approaching 0. In three dimensional Cartesian coordinates, the divergence is defined to be the above scalar function.

Example:

$$\begin{aligned} \mathbf{u} &= \frac{C\mathbf{r}}{r^3} \\ &= \frac{C}{r^2}\hat{\mathbf{r}} \\ &= C\frac{x\mathbf{i} + y\mathbf{j} + z\mathbf{k}}{(x^2+y^2+z^2)^{\frac{3}{2}}} \\ \Rightarrow \frac{\partial u_1}{\partial x} &= \frac{\partial}{\partial x}\frac{Cx}{(x^2+y^2+z^2)^{\frac{3}{2}}} = C\frac{y^2+z^2-2x^2}{(x^2+y^2+z^2)^{\frac{5}{2}}} \end{aligned}$$

and two similar expressions.We can then derive that

$$\nabla \cdot \mathbf{u} = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial u_3}{\partial z} = 0$$

except at $\mathbf{r} = 0$, where $\nabla \cdot \mathbf{u}$ is infinite!

Some physical examples (inverse square law) include:

(i) gravitational field $C = -Gm$, where mass is the source of the field.

(ii) fluid source $C = \frac{V}{4\pi}$, where V is the volume per second injected.

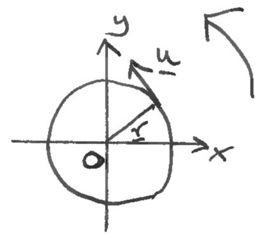


Figure 2.18: solid rotation illustration

(3) $\nabla \times \mathbf{u} \equiv \text{curl } \mathbf{u}$.

It acts as a local rotation. Again, for clear intuition behind the formula, visit the following link to see how 3B1B explains it:

https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/curl-grant-videos/v/2d-curl-formula

Example:

(a) $\mathbf{u} = \mathbf{w} \times \mathbf{r} = w\mathbf{k} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = -wy\mathbf{i} + wx\mathbf{j}$.

$$\Rightarrow \nabla \times \mathbf{u} = 2w\mathbf{k} = 2\mathbf{w}.$$

where $\mathbf{w}$ is a vector pointing upward out of the plane, describing the rate of rotation.

This describes the solid rotation, as illustrated in Figure 2.18, and it turns out the the curl of $\mathbf{u}$ is twice the constant rotation rate $\mathbf{w}$.

(b) $\mathbf{u} = (U + \alpha y, 0, 0)$

This describes uniform shear flow, as illustrated in Figure 2.19. By keeping U and α constant, we can derive that

$$\nabla \times \mathbf{u} = -\alpha\mathbf{k}$$

Note that $\alpha > 0$ in Figure 2.19, resulting in a clockwise rotation as O translates. L $\rightarrow$ R ($y > -\frac{U}{\alpha}$), R $\rightarrow$ L ($y < -\frac{U}{\alpha}$).

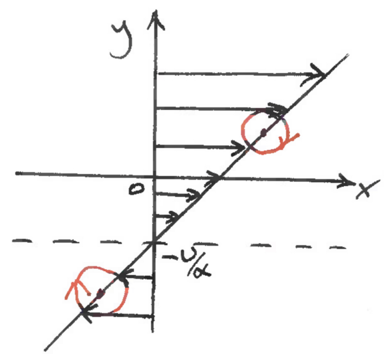


Figure 2.19: uniform shear flow illustration

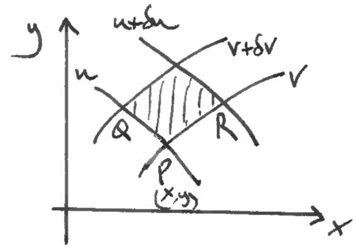


Figure 2.20: Jacobian transform illustration

2.4.10 Application — Double/Repeated Integrals

Solving double integral involves in the change of variable $(x, y) \to (u, v)$, transforming the original double integral into:

$$\iint_A f(x, y)\mathrm{d}x\mathrm{d}y = \iint_A f(x(u, v), y(u, v))J\mathrm{d}u\mathrm{d}v$$

where J is the ***Jacobian*** of the transformation.

One possible illustration is as shown in Figure 2.20. The three critical points are $P(x, y), Q\left(x + \frac{\partial x}{\partial v}\delta v, y + \frac{\partial y}{\partial v}\delta v\right), R\left(x + \frac{\partial x}{\partial u}\delta u, y + \frac{\partial y}{\partial u}\delta u\right)$. Area of the shaded parallelogram is

$$\left|\left(\frac{\partial x}{\partial u}\delta u, \frac{\partial y}{\partial u}\delta u\right) \times \left(\frac{\partial x}{\partial v}\delta v, \frac{\partial y}{\partial v}\delta v\right)\right| = J\delta u\delta v$$

where

$$J = \left| \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \right| = \nabla x(u,v) \times \nabla y(u,v).$$

For example, if Cartesian is transformed to polar, i.e. $(x,y) \to (r,\theta)$, since $x = r\cos\theta, y = r\sin\theta,$ then

$$\frac{\partial x}{\partial r} = \cos\theta,\ \frac{\partial x}{\partial \theta} = -r\sin\theta,\ \frac{\partial y}{\partial r} = \sin\theta,\ \frac{\partial y}{\partial \theta} = r\cos\theta,$$

$$\Rightarrow J = \left| \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \right| = r.$$

2.5 Fourier Integrals

2.5.1 Definitions and Examples

Recap A function $f(x)$ defined on $-L \le x \le L$ (the *fundamental interval*) can be expressed in the form

$$f(x) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L} \right],$$

where

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}\mathrm{d}x,$$
$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi x}{L}\mathrm{d}x,$$

and $n \in \mathbb{Z}$. $f(x)$ is continuous, and the series converges to $f(x)$; at points of discontinuity the series converges to $\frac{1}{2}[f(x_-) + f(x_+)]$. The series is, of course, $2L$ periodic. The complex form is

$$f(x) \sim \sum_{n=-\infty}^{\infty} y_n e^{\frac{in\pi x}{L}},$$

where

$$y_n = \frac{1}{2L}\int_{-L}^{L} f(x) e^{-\frac{in\pi x}{L}}\mathrm{d}x,$$

and $n \in \mathbb{Z}$. We also note the ***Parseval's theorem***:

$$\frac{1}{L}\int_{-L}^{L}(f(x))^2\mathrm{d}x = \frac{1}{2}a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2).$$

We now extend these ideas to that of a function $f(x)$ defined on $(-\infty, \infty)$, by taking the limit $L \to \infty$. The complex form can be combined as

$$f(x) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\left[\int_{-L}^{L} f(s)e^{-i\omega_n s}\mathrm{d}s\right]e^{i\omega_n x}\delta\omega$$

where $\omega_n = \frac{n\pi}{L}$, $\delta\omega = \omega_{n+1} - \omega_n = \frac{\pi}{L}$, and the equality is interpreted as previously at points of continuity/discontinuity of $f(x)$. Formally, we allow $L \to \infty$ so that $\delta\omega \to 0$.

Theorem 101. The ***Fourier transform*** is defined as

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\mathrm{d}x$$

and the ***inverse Fourier transform*** is defined as

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega x}\mathrm{d}\omega.$$

Together, they are called the ***Fourier transform pair***.

Of course the Fourier Transform needs a proof. For now we assume that $\int_{-\infty}^{\infty}|f(x)|\mathrm{d}x$ converges and that $f(x)$, $f'(x)$ are continuous for all x — these requirements may be relaxed later.

We write the RHS of the transform in the form

$$\begin{aligned}
f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f(s)e^{-i\omega s}\mathrm{d}s\right\}e^{i\omega x}\mathrm{d}\omega \\
&= \lim_{L\to\infty}\frac{1}{2\pi}\int_{-L}^{L}\left[\int_{-\infty}^{\infty} f(s)e^{-i\omega(s-x)}\mathrm{d}s\right]\mathrm{d}\omega \\
&= \lim_{L\to\infty}\frac{1}{2\pi}\int_{-L}^{L}\left\{\int_{-\infty}^{\infty} f(s)\cos\left[\omega(s-x)\right]\mathrm{d}s\right. \\
&\qquad\qquad \left. - i\int_{-\infty}^{\infty} f(s)\sin\left[\omega(s-x)\right]\mathrm{d}s\right\}\mathrm{d}\omega
\end{aligned}$$

The first term in $\hat{f}(\omega)$ is even about $\omega = 0$, and the second term is odd. Because the inner integral is absolutely convergent we can change the order of integration. So we get

$$\begin{aligned}
&\lim_{L\to\infty} \frac{1}{\pi} \int_0^L \left\{ \int_{-\infty}^{\infty} f(s)\cos\left[\omega(s-x)\right]\mathrm{d}s \right\} \mathrm{d}\omega \\
&= \lim_{L\to\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f(s) \left\{ \int_0^L \cos\left[\omega(s-x)\right]\mathrm{d}\omega \right\} \mathrm{d}s \\
&= \lim_{L\to\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f(s) \frac{\sin\left[L(s-x)\right]}{s-x}\mathrm{d}s \\
&= \lim_{L\to\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f(x+u) \frac{\sin Lu}{u}\mathrm{d}u \qquad \text{(substitute } s = x+u\text{)} \\
&= \lim_{L\to\infty} \frac{1}{\pi} \left\{ \int_{-\infty}^{\infty} \frac{f(x+u)-f(x)}{u} \sin Lu\mathrm{d}u + f(x)\int_{-\infty}^{\infty} \frac{\sin Lu}{u}\mathrm{d}u \right\}.
\end{aligned}$$

The first integral tends to zero as $L \to \infty$ by the Riemann-Lebesgue lemma. We then note that $\int_{-\infty}^{\infty} \frac{\sin p}{p}\mathrm{d}p = \pi$. We then, finally, get $f(x)$ for this limit, and obtain the previously defined Fourier transform pair.

Example 102. Rectangular wave

$$f(x) = \begin{cases} 1 & |x| \le d \\ 0 & |x| \ge d. \end{cases}$$

$$\begin{aligned}
\hat{f}(\omega) &= \int_{-d}^{d} 1 \cdot e^{-i\omega x}\mathrm{d}x \\
&= \left[\frac{e^{-i\omega x}}{-i\omega} \right]_{-d}^{d} \\
&= -\frac{1}{i\omega}(e^{-i\omega d} - e^{i\omega d}) \\
&= \frac{2}{\omega}\sin\omega d.
\end{aligned}$$

We note that at a point $x = x_0$ of discontinuity of $f(x)$, $f(x)$ in the Fourier transform becomes $\frac{1}{2}(f(x_0^-) + f(x_0^+))$.

2.5.2 Cosine and Sine Transforms

Symmetry was exploited to define half range Fourier *series*. We can now do the same thing for *transforms* over the range $[0, \infty)$.If $f(x)$ is even about $x = 0$ then we can write

$$\begin{aligned}\hat{f}(\omega) &= \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(x)(\cos\omega x - i\sin\omega x)\mathrm{d}x \\ &= 2\int_{0}^{\infty} f(x)\cos\omega x\mathrm{d}x.\end{aligned}$$

We define

$$\hat{f}_c(\omega) = \int_0^{\infty} f(x)\cos\omega x\mathrm{d}x.$$

So

$$\hat{f}(\omega) = 2\hat{f}_c(\omega).$$

Of course $\hat{f}_c(\omega)$ is even about $\omega = 0$, implying that

$$\begin{aligned}f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega x}\mathrm{d}\omega \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \hat{f}_c(\omega)e^{i\omega x}\mathrm{d}\omega \\ &= \frac{2}{\pi}\int_{0}^{\infty} \hat{f}_c(\omega)\cos\omega x\mathrm{d}\omega.\end{aligned}$$

Similarly, if $f(x)$ is odd about $x = 0$ we can define the Fourier sine transform

$$\hat{f}_s(\omega) = \int_0^{\infty} f(x)\sin\omega x\mathrm{d}x$$

so that

$$f(x) = \frac{2}{\pi}\int_0^{\infty} \hat{f}_s(\omega)\sin\omega x\mathrm{d}\omega.$$

2.5.3 Properties of Fourier Transforms

(i) Linearity:

$$h(x) = af(x) + bg(x) \longleftrightarrow \hat{h}(\omega) = a\hat{f}(\omega) + b\hat{g}(\omega)$$

where a, b are constants.

Proof. Exercise! □

(ii) Time scaling:

$$a \in \mathbb{R}, a \neq 0 \Longrightarrow \widehat{f(ax)} = \frac{1}{|a|}\hat{f}\left(\frac{\omega}{a}\right).$$

Alternatively, if $h(x) = f(ax)$, then

$$\hat{h}(\omega) = \frac{1}{|a|}\hat{f}\left(\frac{\omega}{a}\right).$$

Proof.

$$\begin{aligned}
\hat{f}(\omega) &= \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\mathrm{d}x \\
&= \int_{-\infty}^{\infty} f(ax)e^{-i\omega ax}\mathrm{d}(ax) \\
&= |a|\int_{-\infty}^{\infty} f(ax)e^{-i\omega ax}\mathrm{d}x.
\end{aligned}$$

By additional substitution on ω, we get

$$\frac{1}{|a|}\hat{f}\left(\frac{\omega}{a}\right) = \int_{-\infty}^{\infty} f(ax)e^{-i\omega x}\mathrm{d}x.$$

Alternative approach:

$$\begin{aligned}
f(ax) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega ax}\mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{|a|}\hat{f}(\omega)e^{i\omega ax}d(a\omega) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{|a|}\hat{f}\left(\frac{\omega}{a}\right)e^{i\omega x}\mathrm{d}\omega \\
&= h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{h}(\omega)e^{i\omega x}\mathrm{d}\omega
\end{aligned}$$

And the final most direct, intuitive approach:

$$\begin{aligned}\widehat{f(ax)} &= \int_{-\infty}^{\infty} f(ax)e^{-i\omega x}\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(y)e^{-i\frac{\omega}{a}y}\mathrm{d}\left(\frac{y}{a}\right) \\ &= \frac{1}{|a|}\int_{-\infty}^{\infty} f(y)e^{-i\frac{\omega}{a}y}\mathrm{d}y \\ &= \frac{1}{|a|}\hat{f}\left(\frac{\omega}{a}\right).\end{aligned}$$

□

Specially, $\widehat{f(-x)} = \hat{f}(-\omega)$, i.e. $h(x) = f(-x) \Rightarrow \hat{h}(\omega) = \hat{f}(-\omega)$, the so-called *time-reversal* property.

(iii) Translation/time shifting:

$$\forall x_0 \in \mathbb{R}, f\widehat{(x - x_0)} = e^{-i\omega x_0}\hat{f}(\omega).$$

Alternatively, if $h(x) = f(x - x_0)$, then

$$\hat{h}(\omega) = e^{-i\omega x_0}\hat{f}(\omega).$$

Proof.

$$\begin{aligned}\hat{f}(\omega) &= \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(x - x_0)e^{i\omega(x_0 - x)}\mathrm{d}(x - x_0)\end{aligned}$$

By multiplying $e^{-i\omega x_0}$, we obtain

$$e^{-i\omega x_0}\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x - x_0)e^{i\omega x}\mathrm{d}x.$$

Alternative approach:

$$\begin{aligned} f(x-x_0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega(x-x_0)}\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega x_0}\hat{f}(\omega)e^{i\omega x}\mathrm{d}\omega \\ &= h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{h}(\omega)e^{i\omega x}\mathrm{d}\omega \end{aligned}$$

The most direct approach:

$$\begin{aligned} \widehat{f(x-x_0)} &= \int_{-\infty}^{\infty} f(x-x_0)e^{-i\omega x}\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(y)e^{-i\omega(y+x_0)}\mathrm{d}(y+x_0) \\ &= (e^{-i\omega x_0})\int_{-\infty}^{\infty} f(y)e^{-i\omega y}\mathrm{d}y \\ &= e^{-i\omega x_0}\hat{f}(x). \end{aligned}$$

□

(iv) Modulation/frequency shifting:

$$\forall \omega_0 \in \mathbb{R}, \widehat{e^{i\omega_0 x}f(x)} = \hat{f}(\omega-\omega_0).$$

Alternatively, if $h(x) = e^{i\omega_0 x}f(x)$, then

$$\hat{h}(\omega) = \hat{f}(\omega-\omega_0).$$

Proof. Exercise! □

(v) Symmetry:

$$\widehat{\hat{f}(x)} = 2\pi f(-\omega).$$

Proof.

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega x}\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(s)e^{isx}\mathrm{d}s. \end{aligned}$$

And then by another substitution of $x = -\omega$, we get

$$\begin{aligned} f(-\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(s)e^{-i\omega s}\mathrm{d}s \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(x)e^{-i\omega x}\mathrm{d}x \\ &= \frac{1}{2\pi}\widehat{\hat{f}(x)}. \end{aligned}$$

And the direct approach:

$$\begin{aligned} \widehat{\hat{f}(x)} &= \int_{-\infty}^{\infty} \hat{f}(x)e^{-i\omega x}\mathrm{d}x \\ &= 2\pi f(-\omega). \end{aligned}$$

□

(vi) Derivatives: (Useful for ODE, PDE...)

If f and its derivatives $\to 0$ at $\pm\infty$, (i.e. the derivatives have ***compact support***)

$$\begin{aligned} \frac{\widehat{\mathrm{d}^n f(x)}}{\mathrm{d}x^n} &= \int_{-\infty}^{\infty} \frac{\mathrm{d}^n f}{\mathrm{d}x^n}e^{-i\omega x}\mathrm{d}x \\ &= \left[\frac{\mathrm{d}^{n-1} f}{\mathrm{d}x^{n-1}}e^{-i\omega x}\right]_{-\infty}^{\infty} + i\omega\int_{-\infty}^{\infty} \frac{\mathrm{d}^{n-1} f}{\mathrm{d}x^{n-1}}e^{-i\omega x}\mathrm{d}x \\ &\vdots \\ &= (i\omega)^n \hat{f}. \end{aligned}$$

(vii)

$$\begin{aligned} \widehat{xf(x)} &= \int_{-\infty}^{\infty} xf(x)e^{-i\omega x}\mathrm{d}x \\ &= i\frac{\mathrm{d}}{\mathrm{d}\omega}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}\mathrm{d}x \\ &= i\frac{\mathrm{d}}{\mathrm{d}\omega}\widehat{f(\omega)}. \end{aligned}$$

(viii) For cosine and sine transforms we have similarly (for reference only)

(a) $\widehat{f'(x)}_c = -f(0) + \omega \hat{f}_s(\omega)$

Proof.

$$\begin{aligned}\widehat{f'(x)}_c &= \int_0^\infty f'(x)\cos\omega x \mathrm{d}x \\ &= [f(x)\cos\omega x]_0^\infty + \omega\int_0^\infty f(x)\sin\omega x \mathrm{d}x \\ &= -f(0) + \omega\hat{f}_s(\omega)\end{aligned}$$

where $f(\infty) = 0$ due to the "*compact support*" property of derivatives. □

(b) $\widehat{f'(x)}_s = -\omega \hat{f}_c(\omega)$

Proof. Exercise! □

(c) $\widehat{f''(x)}_c = -f'(0) - \omega^2 \hat{f}_c(\omega)$

Proof. Exercise! □

(d) $\widehat{f''(x)}_s = \omega f(0) - \omega^2 \hat{f}_s(\omega)$.

Proof. Exercise! □

(ix) If $f(x)$ is *complex* with conjugate $[f(x)]^*$, then

$$\widehat{[f(x)]^*} = \left[\hat{f}(-\omega)\right]^*.$$

Proof.

$$\widehat{[f(x)]^*} = \int_{-\infty}^\infty [f(x)]^* e^{-i\omega x}\mathrm{d}x.$$

On the other hand,

$$\hat{f}(-\omega) = \int_{-\infty}^\infty f(x)e^{i\omega x}\mathrm{d}x$$

$$\Rightarrow \left[\hat{f}(-\omega)\right]^* = \int_{-\infty}^\infty [f(x)]^* e^{-i\omega x}\mathrm{d}x.$$

□

2.5.4 The Convolution Theorem

Definition 103. For two functions $f(x)$, $g(x)$, we define their ***convolution*** to be

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(x-u)g(u)\mathrm{d}u.$$

Proposition 104.

$$\widehat{f(x) * g}(x) = \hat{f}(\omega)\hat{g}(\omega).$$

Proof.

$$\begin{aligned}
\widehat{f(x) * g}(x) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x-u)g(u)\mathrm{d}u\right] e^{-i\omega x}\mathrm{d}x \\
&\quad \text{(change order of integration)} \\
&= \int_{-\infty}^{\infty} g(u) \left[\int_{-\infty}^{\infty} f(x-u)e^{-i\omega x}\mathrm{d}x\right] \mathrm{d}u \\
&\quad (s = x-u) \\
&= \int_{-\infty}^{\infty} g(u) \left[\int_{-\infty}^{\infty} f(s)e^{-i\omega(s+u)}\mathrm{d}s\right] \mathrm{d}u \\
&= \left(\int_{-\infty}^{\infty} g(u)e^{-i\omega u}\mathrm{d}u\right)\left(\int_{-\infty}^{\infty} f(s)e^{-i\omega s}\mathrm{d}s\right) \\
&= \hat{f}(\omega)\hat{g}(\omega).
\end{aligned}$$

□

Example 105. Consider $\hat{f}(\omega) = \frac{1}{4+\omega^2}$, $\hat{g}(\omega) = \frac{1}{9+\omega^2}$. It can be shown that $f(x) = \frac{1}{4}e^{-2|x|}$, $g(x) = \frac{1}{6}e^{-3|x|}$, so that the inverse transform of $\frac{1}{(4+\omega^2)(9+\omega^2)}$ is

$$\begin{aligned}
f * g &= \frac{1}{24}\int_{-\infty}^{\infty} e^{-2|x-u|}e^{-3|u|}\mathrm{d}u \\
&\ \ \vdots \\
&= \frac{1}{20}e^{-2|x|} - \frac{1}{30}e^{-3|x|}.
\end{aligned}$$

2.5.5 The Plancherel/Energy Theorem

As we should expect, there is an analogue to Parseval's theorem for Fourier series.

Theorem 106. If $f(x)$ is a real valued function, then

$$\int_{-\infty}^{\infty} [f(u)]^2 \mathrm{d}u = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\hat{f}(\omega)\right|^2 \mathrm{d}\omega.$$

Proof. We can use the previous properties to show that

$$[\widehat{f(-x)}]^* = [\hat{f}(\omega)]^*$$

and apply this for real $f(x)$ we get

$$\widehat{f(-x)} = [\hat{f}(\omega)]^*.$$

Now we use the convolution theorem with $\hat{g}(\omega) = [\hat{f}(\omega)]^*$, we get

$$f(x\widehat{) * f(}-x) = \hat{f}(\omega)[\hat{f}(\omega)]^* = \left|\hat{f}(\omega)\right|^2.$$

So inverting we get

$$f(x) * f(-x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\hat{f}(\omega)\right|^2 e^{i\omega x} \mathrm{d}\omega.$$

The LHS of the above is equation is $= \int_{-\infty}^{\infty} f(u+x)f(u)\mathrm{d}u$. Put $x = 0$ to obtain the result. □

2.5.6 The Dirac Delta Function

Definition 107. We define the ***Dirac Delta*** function as

$$\delta(x) = \lim_{k\to\infty} f_k(x).$$

where

$$f_k(x) = \begin{cases} \frac{k}{2} & (|x| < \frac{1}{k}) \\ 0 & (|x| > \frac{1}{k}) \end{cases}$$

Of course,

$$\int_{-\infty}^{\infty} f_k(x)\mathrm{d}x = 1.$$

Note This is *not* a function in the normal sense — it is a *generalised function*, being ∞ at $x = 0$ and 0 for $x \neq 0$ with $\int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1$.

Shifting property

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x)\delta(x)\mathrm{d}x &= \lim_{k\to\infty} \int_{-\infty}^{\infty} g(x) f_k(x)\mathrm{d}x \\
&= \lim_{k\to\infty} \int_{-\frac{1}{k}}^{\frac{1}{k}} g(x)\frac{k}{2}\mathrm{d}x \\
&= \lim_{k\to\infty} \left[\frac{k}{2}\cdot\frac{2}{k} g(\overline{x})\right] \\
&= g(0)
\end{aligned}$$

where the second last step is by the mean value theorem, that $-\frac{1}{k} < \overline{x} < \frac{1}{k}$. This means that

$$\int_{-\infty}^{\infty} g(x)\delta(x-a)\mathrm{d}x = g(a).$$

We can also derive that

$$\widehat{\delta(x)} = \int_{-\infty}^{\infty} e^{-i\omega x}\delta(x)\mathrm{d}x = e^{-i\omega 0} = 1$$

so that $\delta(x)$ is the inverse transform of 1. Naturally we can then write

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\pm i\omega x}\mathrm{d}\omega$$

as an alternative representation, and, of course,

$$\delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\pm i\omega x}\mathrm{d}x.$$

($\pm$ is because $\delta(x) = \delta(-x)$.)

So this helps us to find Fourier transforms of functions that do not decay as $x \to \pm\infty$.

Example 108. Find the fourier transform of $\cos\omega_0 x$.

$$\begin{aligned}
\widehat{\cos\omega_0 x} &= \int_{-\infty}^{\infty} \frac{1}{2}(e^{i\omega_0 x} + e^{-i\omega_0 x})e^{-i\omega x}\mathrm{d}x \\
&= \frac{1}{2}\int_{-\infty}^{\infty} e^{-i(\omega-\omega_0)x}\mathrm{d}x + \frac{1}{2}\int_{-\infty}^{\infty} e^{-i(\omega+\omega_0)x}\mathrm{d}x \\
&= \pi\delta(\omega-\omega_0) + \pi\delta(\omega+\omega_0).
\end{aligned}$$

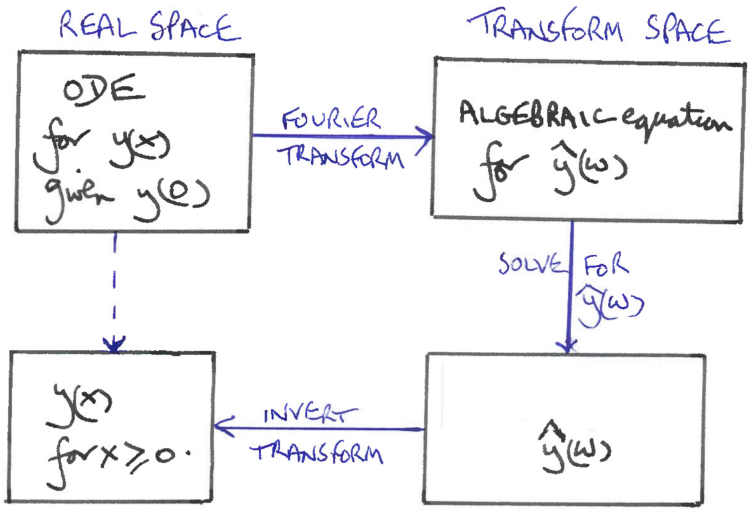


Figure 2.21: Fourier transform application

Exercise: Find the fourier transform of $\sin x$.

2.5.7 Application of Transforms — To Come!

A general principle in solving e.g. a differential equation would be to transform from real space to transform space, solve these and then invert back, as shown in Figure 2.21.

Example 109.

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + y = f(x)$$

From the Fourier transform property for derivatives, we can transform the first term

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = (i\omega)^2\hat{y} = -\omega^2\hat{y}$$

following the other terms, deriving

$$\hat{y} = \frac{\hat{f}}{1-\omega^2}$$

and thus,

$$\Rightarrow y(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\hat{f}}{1-\omega^2}e^{i\omega x}\mathrm{d}\omega.$$

If $f(x)$ is 'simple' then we can use the transform properties to perform this — if not then more integration technique is required!

Chapter 3

Linear Algebra

3.1 Introduction to Matrices and Vectors

3.1.1 Column vectors

Definition 110. A *column vector* (n-column vector) $\mathbf{v}_n$ is a tuple of n real numbers written as a single columnn, with $a_1, a_2, a_3, \ldots, a_n \in \mathbb{R}$:

$$\mathbf{v}_n := \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix}$$

Definition 111. $\mathbb{R}^n$ is the set of all column vectors of height n whose entries are real numbers. In symbols:

$$\mathbb{R}^n = \{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} : a_1, a_2, \ldots, a_n \in \mathbb{R}\}$$

Example 112. $\mathbb{R}^2$ can be seen as Euclidean plane. $\mathbb{R}^3$ can be seen as Euclidean space.
<u>Caution</u>: Our vectors always “start” at the origin.

Definition 113. The ***zero vector*** $\mathbf{0}_n$ is the height n-column vector all of whose entries are 0.

Definition 114. The ***standard basis vectors*** in $\mathbb{R}^n$ are the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

i.e. $\mathbf{e}_k$ is the vector with kth entry equal to 1 and all other entries equal to 0.

Operations on column vectors

$$\mathbf{v} := \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \quad \mathbf{u} := \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

be column vectors $\mathbb{R}^n$, and let λ be a (real or complex) number.

(1) Addition on vectors in $\mathbb{R}^n$ is given by:

$$\begin{pmatrix} v_1 + u_1 \\ v_2 + u_2 \\ \vdots \\ v_n + u_n \end{pmatrix}$$

$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ (binary operatioin). $(\mathbb{R}^n, +)$ is a group.

(2) ***Scalar multiplication*** $\lambda\mathbf{v}$ on $\mathbb{R}^n$:

$$\begin{pmatrix} \lambda v_1 \\ \lambda v_2 \\ \vdots \\ \lambda v_n \end{pmatrix}$$

$s : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, so not binary operation.

(3) ***Dot product*** $v \cdot u$ is defined to be the number $v_1u_1 + v_2u_2 + \cdots + v_nu_n$. $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, so not binary.

Example 115. Show that $(\mathbb{R}^n, +)$ is an Abelian group.

- Identity: $\mathbf{0}_n$ $(v + \mathbf{0}_n = \mathbf{v})$
- $-\mathbf{v}$ are inverses, where

$$-\mathbf{v} := \begin{pmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{pmatrix}$$

- associativity: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
- commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

<u>Caution</u>: $+$ only makes sense for vectors of the *same size.* e.g. $\mathbf{v} \cdot \mathbf{0}_n = 0 \in \mathbb{R}$.

Definition 116. let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n \in \mathbb{R}^n, \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n \in \mathbb{R}$, then

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n$$

is called a ***linear combination*** of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n$.

Definition 117. The set of all linear combinatioins of a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is called the ***span*** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.
Notation:
$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} := \{\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n | \lambda_1, \ldots, \lambda_n \in \mathbb{R}\}$

Example 118. compute the span of

- $\{\mathbf{e}_1, \mathbf{e}_2\}$, $\mathbf{e}_1, \mathbf{e}_2 \in \mathbb{R}^2$.

$$\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \{\lambda_1\mathbf{e}_1 + \lambda_2\mathbf{e}_2 | \lambda_1, \lambda_2 \in \mathbb{R}\} = \{\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} | \lambda_1, \lambda_2 \in \mathbb{R}\}$$

- $\text{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}\right\} = \{\begin{pmatrix} \lambda_1 \\ 2\lambda_2 \\ 0 \end{pmatrix} | \lambda_1, \lambda_2 \in \mathbb{R}\}$

Definition 119. let $\mathbf{v} \in \mathbb{R}^n$. The ***length*** of $\mathbf{v}$, a.k.a. the ***norm*** of $\mathbf{v}$, is the non-negative real number $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Note: $\|\mathbf{0}\| = 0$, and conversely if $\mathbf{v} \neq 0$ then $\|\mathbf{v}\| > 0$. This definition agrees with out usual ideas about the length of a vector in $\mathbb{R}^2$ or $\mathbb{R}^3$, which follows from Pythagoras' theorem.

Definition 120. A vector $\mathbf{v} \in \mathbb{R}^n$ is called a ***unit vector*** if $\|\mathbf{v}\| = 1$.

Example 121.

(1) Any non-zero vector $\mathbf{v}$ can be made into a unit vector $\hat{\mathbf{u}} := \frac{\mathbf{v}}{\|\mathbf{v}\|}$. This process is called ***normalizing***.

(2) The standard basis vectors are unit vectors.

3.1.2 Basic Matrix Operations

Definition 122. An $n \times m$-matrix is a rectangular grid of numbers called the *entries* of the matrix with n rows and m columns. A real matrix is onne whose entries are real numbers, and a complex matrix is one whose entries are complex numbers.

Notations: $M_{n\times m}(\mathbb{R}), M_{n,m}(\mathbb{R}), \mathrm{Mat}_{n\times m}(\mathbb{R}), \mathbb{R}^{n\times m}$.

Operations on matrices:

Definition 123. let $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times m$-matrix, $\lambda \in \mathbb{R}$. Then:

(1) $A + B = n \times m$-matrix $(a_{ij} + b_{ij})$. $+ : M_{n\times m}(\mathbb{R}) \times M_{n\times m}(\mathbb{R}) \to M_{n\times m}(\mathbb{R})$

(2) $\lambda A = n \times m$-matrix (λa_{ij})

Theorem 124. $(M_{n\times m}(\mathbb{R}), +)$ is an Abelian grouop.

Definition 125. The ***transpose*** A^T of an $n\times m$-matrix (a_{ij}) is the $m\times n$-matrix (a_{ij}). The ***leading diagonal*** of a matrix is the $(1,1),(2,2),\dots$ entries. So the transpose is obtained by doing a reflection in the leading diagonal.

(Multiplying matrices with vectors) Definition 126. Let $A=(a_{ij})$ be an $n\times m$-matrix, $\mathbf{v}\in\mathbb{R}^m$. Then $A\mathbf{v}$ is the vector in $\mathbb{R}^n$ with i-th row entry $\sum_{j=1}^m a_{ij}\mathbf{v}_j$

Example 127.

- Prove that for $A\in M_{n\times m}(\mathbb{R}), \mathbf{e}_k\in\mathbb{R}^m$, $A\mathbf{e}_k = k$-th column of A.

  Proof: let $A=(a_{ij})$. By definition the i-th entry of $A\mathbf{e}_k$ is

  $$\sum_{j=1}^m a_{ij}(\mathbf{e}_k)_j = a_{ik}$$

  since $(\mathbf{e}_k)_j = 0$ whenever $j\neq k$, 1 for $j=k$

- Let I_n be the identity matrix. Show formally that $I_n\nu=\nu$, $\forall \nu\in\mathbb{R}^n$.

- $\nu\cdot\mathbf{v} = \nu^T\mathbf{v}$

- let $\nu_1,\nu_2,\nu_3\in\mathbb{R}^3$. Write the linear combination $3\nu_1-5\nu_2+7\nu_3$ as a multiplication of matrix $A\in M_{3\times 3}(\mathbb{R})$ with a vector $\mathbf{x}\in\mathbb{R}^3$. Then

  $$A\mathbf{x} = \begin{pmatrix}\nu_1 & \nu_2 & \nu_3\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = x_1\nu_1+x_2\nu_2+x_3\nu_3$$

  with ν_1,ν_2,ν_3 written as a column vector to form a matrix in the above expression, thus using matrix multiplication to express linear combination of vectors.

3.2 Systems of linear equations

3.2.1 Definitions

Definition 128. A ***linear equation*** in the variables $x_1, x_2, \dots, x_n \in \mathbb{R}$ is an equation of the form:

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n = c, \text{ with } \lambda_1, \dots, \lambda_n \subset \textit{ Fixed} \text{ real numbers}$$

Caution: In particular, no powers/multiplications/function of one or more variables.

Definition 129. A system of n linear equations is a list of simultaneous linear equations. It can be converted to $A\mathbf{x} = \mathbf{b} \in \mathbb{R}^m$,with

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n}$$

Caution: Thee $m \times n$-matrix A is called coefficient matrix. The matrix $(A|\mathbf{b})$ where the vector $\mathbf{b}$ is added as a column on the right is called ***augmented matrix***.

Definition 130. A system is called ***consistent*** (resp. inconsistent) if it has a solution $(s_1, s_2, \dots, s_m)$ (resp. no solution).

Example 131.

$$\begin{cases} x_1 + x_3 - x_4 = 1 \\ x_2 - x_4 = 6 \\ x_1 + x_2 + 6x_3 - 3x_4 = 0 \end{cases}$$

Augmented matrix form:

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -1 & 1 \\ 0 & 1 & 0 & -1 & 6 \\ 1 & 1 & 6 & -3 & 0 \end{array}\right)$$

Definition 132. A ***row operation*** is one of the following procedures on a $n \times m$-matrix (a_{ij}):

(1) $r_i(\lambda)$: multiply row i by a scalar $\lambda \in \mathbb{R}, \lambda \neq 0$.

(2) r_{ij}: swap row i with row j.

(3) $r_{ij}(\lambda)$: multiply row i by $\lambda \neq 0$, $\lambda \in \mathbb{R}$ and add it to row j.

Example 133. let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, so

$$r_{12} \Rightarrow \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$$

$$r_2(2) \Rightarrow \begin{pmatrix} 1 & 2 \\ 6 & 8 \end{pmatrix}$$

$$r_{12}(2) \Rightarrow \begin{pmatrix} 1 & 2 \\ 5 & 8 \end{pmatrix}$$

Proposition 134. Let $A\mathbf{x} = \mathbf{b}$ be a system of linear equations in matrix form, $(A|\mathbf{b})$ the augmented matrix, $(A'|\mathbf{b}')$ the augmented matrix of the system after row operation. Show that x is solution of $A\mathbf{x} = \mathbf{b} \iff x$ is solution of $A'\mathbf{x} = \mathbf{b}'$.

Proof. row operations of type (1) and (2) $\Rightarrow$ trivial.

(3) Take equation i, multiply it by λ, add it to equation j. $\Rightarrow (a_{j1} + \lambda a_{i1})x_1 + \cdots + (a_{jm} + \lambda a_{im})x_m = b_j + \lambda b_i$. $\square$

Caution: Every row operation is invertible:

$$[r_i(\lambda)]^{-1} = r_i(\frac{1}{\lambda}),\ [r_{ij}]^{-1} = r_{ij},\ [r_{ij}(\lambda)]^{-1} = r_{ij}(-\lambda)$$

3.2.2 Gauss algorithm

Definition 135. The left most non-zero entry in a non-zero row is called ***leading entry***. A matrix is called in ***echelon form*** if:

(1) The leading entry in each non-zero row is 1.

(2) The leading 1 of each row is to *the right* of the leading 1 in the row above.

(3) The zero-rows are *below* all other rows.

Example 136.

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Only the last one is in echelon form.

Definition 137. A matrix is ***row reduced echelon form*** if:

(1) It is in echelon form.

(2) The leading 1 in each row is the *only* non-zero entry in its column.

Example 138.

$$\begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \alpha & \beta & 2 \\ 0 & 0 & 1 & -2 \end{pmatrix}.$$

The second one is not, unless $\beta = 0$.

The point of RRE form is that if we have a system of equations

$$A\mathbf{x} = \mathbf{b}$$

and A is in RRE form, then we can easily read off the solution (if any). There are four cases to consider:

(1) Every column of A contains a leading 1, and there are no zeros row. In this case the only possibility is that $A = I_n$ is the identity matrix.

Then the equations are simply

$$\begin{aligned} x_1 &= b_1 \\ x_2 &= b_2 \\ &\vdots \\ x_n &= b_n \end{aligned}$$

and they have a unique solution, the entries of $\mathbf{b}$.

(2) Every column of A contains a leading 1, and there are some zero rows. Then A must have more rows than columns, and it must be a matrix of the form

$$A = \begin{pmatrix} I_n \\ \mathbf{0}_{k\times n} \end{pmatrix}$$

i.e. it looks like an identity matrix with a block of zeros underneath. In this case, the first n equations are

$$\begin{aligned} x_1 &= b_1 \\ x_2 &= b_2 \\ &\vdots \\ x_n &= b_n \end{aligned}$$

and the last k equations are

$$\begin{aligned} 0 &= b_{n+1} \\ 0 &= b_{n+2} \\ &\vdots \\ 0 &= b_{n+k} \end{aligned}$$

Now there are two possibilities:

- If any of the last k entries of $\mathbf{b}$ are non-zero then this system has no solutions, because the last k equations are never satisfied for any $\mathbf{x}$ and the system is inconsistent.
- If the last k entries of $\mathbf{b}$ are all zero then the system has a unique solution, given by setting $x_i = b_i$ for each $i \in [1, n]$.

(3) Some columns of A do not contain a leading 1, but there are no zero rows, for instance

$$A = \begin{pmatrix} 1 & 0 & a_{13} \\ 0 & 1 & a_{23} \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} 1 & a_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If the ith column of A does not contain a leading 1 then the corresponding variable x_i is called a ***free variable***, or free parameter. These variables can be set to any values. Each remaining variable is called a ***basic variable*** and we have a single equation

$$x_j + (\ldots) = b_j$$

where the expression in the brackets only contains free parameters. This equations determines the value of x_j, in terms of the entries in $\mathbf{b}$ and the values of the free parameters. This kind of system always has infinitely many solutions, we say it is ***underdetermined***.

Definition 139. A leading entry in a matrix in RRE form is also called a ***Pivot position***. A ***Pivot column*** is a column containing a Pivot position.

(Gauβ algorithm) Proposition 140. Any matrix can be put into RRE form by performing a sequence of row operations.

Proof. Our proof will consist of the explicit description of the algorithm. Let A be an arbitrary matrix. Step 1—Step 3 below is called the ***forward phase*** and is used to bring the matrix A into echelon form. Step 4 is called the *backwad phase* and is used to bring A inito RRE form.

Step 1: Choose your first pivot position, which is the first non-zero leading term. Do row operation such that the leading term becomes 1.

Step 2: Create zeros below your first leading entry by multiplying the row with the leading entry and subtract it from the subsequent rows.

Step 3: Repeat the first two steps to bring the whole matrix into echelon form.

Step 4: Create zeros above the leading entries to convert to RRE row by row, by multiplying the row where the selected leading entry is in, and subtract it from the above rows.

□

It is also true (althouth we won't show this) that the RRE form of a matrix is unique; if you apply any sequence of row operations which puts your matrix into RRE form, the result is the same as the output of the algorithm we just described.

Now we have a systematic procedure for solving a system of simultaneous linear equations $A\mathbf{x} = \mathbf{b}$:

(1) Form the augmented matrix $(A|\mathbf{b})$.

(2) Apply the algorithm above to put the augmented matrix into RRE form $(A'|\mathbf{b}')$.

(3) Read off the solutions to $A'\mathbf{x} = \mathbf{b}'$

In fact it's not necessary to get the whole matrix $(A'|\mathbf{b}')$ into RRE form, you can stop when the left block A' is in RRE form. Doing further operations to adjust the final column will not help you read the solutions.

Example 141. Solve

$$\begin{cases} 3x_1 + 5x_2 - 4x_3 = 0 \\ -3x_1 - 2x_2 + 4x_3 = 0 \\ 6x_2 + x_2 - 8x_3 = 0. \end{cases}$$

The RRE form of the above equation is

$$\begin{pmatrix} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and the geometric interpretation of this is a line!

Proposition 142. The number of solutions to a system $A\mathbf{x} = \mathbf{b}$ is always either 0, 1, or ∞.

Proof. Assume the number of solutions is not 0, and not 1. Take 2 solutions ν and υ, $\nu \neq \upsilon$.

$$\Rightarrow A\nu = A\upsilon = b \Rightarrow A(\nu - \upsilon) = 0$$

Note that $\nu - \upsilon \neq 0$, and let $\omega = \nu - \upsilon$. Take: $\nu + \lambda\omega, \lambda \in \mathbb{R}$

$$\Rightarrow A(\nu + \lambda\omega) = A\nu + \lambda A\omega = A\nu = \mathbf{b} = \mathbf{b}$$

So $\nu + \lambda\omega$ is a solution $\forall \lambda \in \mathbb{R} \Rightarrow \infty$ many solutions. $\square$

3.3 Matrix Multiplication

3.3.1 Basics of Matrix Multiplication

Definition 143. $A \in M_{m,n}(\mathbb{R}), B \in M_{n,k}(\mathbb{R})$. Then the product AB is defined such that the $(AB)_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$ (row i column k)

Operation: $M_{m,n}(\mathbb{R}) \times M_{n,k}(\mathbb{R}) \to M_{m,k}(\mathbb{R})$. It is a binary operation on $M_{n,n}(\mathbb{R})$, square matrices! Be careful with the size of the matrices.

Caution:

- The (i,j)-entry of AB is the dot product of r_i^T with c_j.
- Other way to see it: column j of AB is Ac_j.

Proposition 144. Let $A, A' \in M_{m,n}(\mathbb{R}), B, B' \in M_{n,p}(\mathbb{R})$. Then

(1) $A(BC) = (AB)C$. (Associativity)

(2)
$$\begin{Bmatrix} A(B+B') = AB + AB' \\ (A+A')B = AB + A'B \end{Bmatrix} \quad \text{Distributivity}$$

(3) $\forall \lambda \in \mathbb{R}, (\lambda A)B = A(\lambda B) = \lambda(AB)$. (Compatibility with scalar multiplication.)

Caution:

- Let $A \in M_{m,n}(\mathbb{R})$, then $0_{k\times m}A = 0_{k\times n}, A0_{n\times e} = 0_{m\times e}$.
- $\forall A \in M_{n,n}(\mathbb{R}), I_nA = AI_n = A$.
- In general, $AB \neq BA$, i.e. not commutative.
- A^2 does not guarantee to be $0_{n,n}$, e.g. $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$

Definition 145. A ***diagonal matrix*** is a square matrix $D \in M_{n,n}(\mathbb{R})$, s.t.

$$\begin{cases} D_{ij} = 0, i \neq j \\ D_{ij} = \lambda_i \in \mathbb{R}, i = j \end{cases}$$

Goal: When can we bring matrices to this form? $\rightsquigarrow$ diagonalization. E.g. $I_n, 0_{n \times n}$.

Definition 146. ***Lower triangular matrix***: $a_{ij} = 0, i < j$. (strictly lower if $i \leq j$) ***Upper triangular matrix***: $a_{ij} = 0, i > j$. (strictly upper if $i \geq j$)

Example 147. echelon form, RRE (Upper triangular). Lower + upper triangular = diagonal.

3.3.2 Inverse of a Matrix and Invertibility

Please visit section 3.6 for close link between the concepts introduced here and others such as nullity.

Definition 148. Let $AM_{n,n}(\mathbb{R})$. A $n \times n$-matrix A^{-1} is called ***inverse*** of A if:

$$AA^{-1} = I_n = A^{-1}A.$$

Caution: Not all matrices are *invertible*!

Lemma 149.

(1) If A is invertible, then its inverse is unique.

(2) If A is invertible and either $AB = I_n$ or $BA = I_n$, for some $B \in M_{n \times n}(\mathbb{R})$, then $B = A^{-1}$.

Proof.

(1) See group theory.

(2) $B = I_n B = (A^{-1}A)B = A^{-1}(AB) = A^{-1}I_n = A^{-1}$. Same for BA.

$\square$

Lemma 150. Assume $A, B \in M_{n\times n}(\mathbb{R})$, invertible. Then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. Exercise! $\square$

Lemma 151. Let $A \in M_{n\times n}(\mathbb{R})$.

(1) If $\exists v \in \mathbb{R}^n, v \neq \mathbf{0}$ s.t. $Av = \mathbf{0}$, then A is not invertible.

(extended: same if $v \in \mathbb{R}^{nT}, vA = \mathbf{0}$.)

(2) If $\exists B \in M_{n\times n}(\mathbb{R}), B \neq 0_{n\times n}$ s.t. $AB = 0_{n\times n}$ or $BA = 0_{n\times n}$, then A is not invertible.

Proof.

(1) (contrapositive) Assume A is invertible $\iff \exists A^{-1} \in M_{n\times n}(\mathbb{R})$, s.t. $AA^{-1} = A^{-1}A = I_n$. Assume $Av = 0$, so $A^{-1}Av = I_n v = v$, contradiction.

(2) Assume $BA = 0_{n\times n}$. A invertible $\Rightarrow \exists A^{-1}$ s.t. $AA^{-1} = I_n$. Therefore $B = BI_n = B(AA^{-1}) = (BA)A^{-1} = 0_{n\times n}$, contradiction.

$\square$

Corollary 152. If $A \in M_{n\times n}(\mathbb{R})$ invertible, then it cannot have a row/column of zeros.

Proof. Exercise! $\square$

Definition 153. An ***elementary matrix*** R is a matrix which differs from $I_n \in M_{n\times n}(\mathbb{R})$ by only *one* elementary row operation. Multiplying a matrix by the elementary matrix on the left is equivalent of performing an elementary row operation on that matrix.

Type 1 Apply $r_i(\lambda)$ to I_n. Multiplying $A \in M_{n\times n}(\mathbb{R})$ by $R_i(\lambda)$ on the left.

Type 2 Apply r_{ij} to I_n. Multiplying A by R_{ij} on the left swaps row i and j in A.

Type 3 Apply $r_{ij}(\lambda)$ to I_n.

Fact: Elementary matrices are *invertible* since row operations are reversible. You can always produce an inverse of an elementary matrix which represents the reversed process of elementary row operation.

Lemma 154. Let $A \in M_{n\times n}(\mathbb{R})$, A' obtained from A by elementary row operations. Then A invertible $\iff$ A' invertible.

Proof. $A \in M_{n\times n}(\mathbb{R})$. Say we got A' from A by row operation. $A' = RA$ for some R elementary matrix, Both A and R are invertible. Then A' is invertible, and $A'^{-1} = A^{-1}R^{-1}$.

For '$\Leftarrow$', $A' = RA \iff R^{-1}A' = A$. □

Lemma 155. $A \in M_{n\times n}(\mathbb{R})$, A' the RRE form of A. Then: A' invertible $\iff$ A' has no zero rows.

Proof. "$\Leftarrow$": Assume A' has no zero rows $\Rightarrow$ A' has a leading one in each column. $\Rightarrow A' = I_n$. □

Corollary 156. A is invertible $\iff$ Its RRE form is the identity matrix.

Proposition 157. An $n \times n$ matrix A is invertible $\iff$ $\neg\exists \mathbf{v} \neq \mathbf{0}$, $\mathbf{v} \in \mathbb{R}^n$ s.t. $A\mathbf{v} = \mathbf{0}$.

Proof. Exercise! □

Algorithm to compute inverse

Step 1 Write the augmented matrix $(A|I_n)$.

Step 2 Bring $(A|I_n)$ to RRE form $\to$ $(A|I_n)$ $\to$ $(R_1A|R_1I_n)$ $\to$ $\cdots$ $\to$ $(R_m\cdots R_2R_1A|R_m\cdots R_2R_1)$.

Step 3 Read result.

Example 158.

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{pmatrix}$$

and calculate its inverse!

$$\begin{aligned}
(A|I_3) &= \begin{pmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{pmatrix} \\
&\vdots \\
&= \begin{pmatrix} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{pmatrix}
\end{aligned}$$

Caution: If at any point you get a row of 0s, Stop! (matrix not invertible)
Caution: Now that we know how to get inverses. Consider: $A\mathbf{x} = \mathbf{b}$. (system of linear equations) If A invertible: $A^{-1}\mathbf{b} = \mathbf{x}$. (Solution is unique.)

3.3.3 Determinant

Definition 159. let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, so $\det A = a_{11}a_{22} - a_{21}a_{12} \neq 0 \iff A$ is invertible.

Definition 160. Let $A \in M_{n\times n}(\mathbb{R})$. A_{ij} is the ***submatrix*** obtained by deleting row i and column j of A. A_{ij} is called the (i, j)-minor.

Definition 161. The ***determinant*** of an $n \times n$-matrix $A = (a_{ij})$ is given by:

$$\det A = \sum_j (-1)^{j+i} a_{ij} \det A_{ij} = \sum_i (-1)^{i+j} a_{ij} \det A_{ij}.$$

The first expression is called the *expansion along the i-th row*, while the second expression is called the *expansion along the j-th column.*

Example 162. Find the determinant of

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 2 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix}.$$

Rules for determinants Let $A \in M_{n\times n}(\mathbb{R})$.

(1) Operation $r_i(\lambda)$ produces matrix B, s.t. $\det B = \lambda \det A$.

(2) Operation r_{ij} produces matrix B, s.t. $\det B = -\det A$.

(3) Operation $r_{ij}(\lambda)$ leaves determinant invariant.

(4) let $B \in M_{n\times n}(\mathbb{R})$, $\det AB = \det A \cdot \det B$.

(5) $\det A = \det A^T$.

Proposition 163. $n \times n$-matrix A is invertible $\iff \det A \neq 0$.

Proof.

- "$\Rightarrow$": Assume A invertible $\Rightarrow \exists A^{-1}$ s.t. $AA^{-1} = I_n$. ($\det I_n = 1$!) So $\det(AA^{-1}) = \det A \cdot \det A^{-1} = \det I_n = 1$. Therefore $\det A \neq 0 \neq \det A^{-1}$.
- "$\Leftarrow$": Assume $\det A \neq 0$. Since elementary matrices are invertible, their determinants are $\neq 0$, so the determinant of RRE form A' is going to be $\neq 0 \Rightarrow A' = I_n \Rightarrow A$ invertible.

□

Definition 164. Let $\sigma \in S_n$. An ***inversion*** (i, j) of σ is a pair of integers (i, j) s.t.:

$$0 < i < j \leq n \quad \text{and} \quad \sigma(i) \geq \sigma(j).$$

The ***sign*** of a permutation $\operatorname{sgn} \sigma$ is given by

- $+1$, if number of inversion is even
- -1, if number of inversion is odd.

Definition 165. Let $A \in M_{n,n}(\mathbb{R})$, then

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn} \sigma \, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Example 166. Let $A = I_n$. Show that $\det I_n = 1$.

3.4 Eigenvalues and Eigenvectors

3.4.1 Basic Definitions

Definition 167. Let $A \in M_{n\times n}(\mathbb{R}), \lambda \in \mathbb{R}, \lambda$ is an ***eigenvalue*** to the ***eigenvector*** $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ if

$$A\mathbf{v} = \lambda \mathbf{v}.$$

Geometrically, applying A to $\mathbf{v}$ just "rescales" $\mathbf{v}$.

Warning:

- $\mathbf{v}$ eigenvector $\iff (A - \lambda I_n)\mathbf{v} = \mathbf{0}$.
- $I_n \mathbf{v} = \mathbf{v}\, \forall \mathbf{v} \neq \mathbf{0} \Rightarrow$ 1 is the only eigenvalue to *all* vectors.
- $0_{n\times n}\mathbf{v} = \mathbf{0}_n = 0 \cdot \mathbf{v} \Rightarrow$ 0 is the only eigenvalue to *all* vectors.

Proposition 168. Let $A \in M_{n\times n}(\mathbb{R})$. Then λ is eigenvalue of $A \iff A - \lambda I_n$ is not invertible.

Proof. A is not invertible $\iff \exists$ non-zero vector $\mathbf{v}$ s.t. $A\mathbf{v} = \mathbf{0}$. (by Proposition 157) Since we want *non-zero* solutions of $(A - \lambda I_n)\upsilon = 0$, $A - \lambda I_n$ cannot be invertible! $\square$

Example 169. $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. Find the eigenvalues.

Augmented matrix:

$$\begin{pmatrix} 1-\lambda & 2 & 0 \\ 2 & 1-\lambda & 0 \end{pmatrix}.$$

Case 1 $\lambda = 1 \Rightarrow$ RRE is I_2, not what we want!

Case 2 Assume $\lambda \neq 1$. Convert to RRE, and to ensure that the matrix is not invertible, $\lambda^2 - 2\lambda - 3 = 0 \Rightarrow \lambda = -1, 3$.

Definition 170. ***Trace*** of $A_{n\times n}(\mathbb{R})$ tr A is $\sum_{i=1}^n a_{ii}$.

Definition 171. The ***characteristic polynomial*** of $A_{2\times 2}(\mathbb{R})$ is

$$\lambda^2 - \operatorname{tr} A\, \lambda + \det A.$$

In general, for $A_{n\times n}(\mathbb{R})$ is

$$\det(A - \lambda I_n).$$

3.4.2 Diagonalization

Definition 172. Let $A, B \in M_{n\times n}(\mathbb{R})$. A and B are called ***similar*** if $\exists P \in M_{n\times n}(\mathbb{R})$ invertible s.t.

$$A = P^{-1}BP.$$

A is called ***diagonalizable*** if B is diagonal. If A is similar to B, then B is similar to A.

Proposition 173. Similar matrices have the same eigenvalues.

Proof. To solve:

$$\begin{aligned}
0 &= \det(A - \lambda I_n) \\
&= \det(P^{-1}BP - \lambda I_n) \\
&= \det(P^{-1}BP - \lambda P^{-1}P) \\
&= \det(P^{-1}(B - \lambda I_n)P) \\
&= \det P^{-1} \det(B - \lambda I_n) \det P \\
&= \det(P^{-1}P) \det(B - \lambda I_n) \\
&= 1 \cdot \det(B - \lambda I_n) \\
&= \det(B - \lambda I_n).
\end{aligned}$$

□

Warning: The eigenvectors are in general *not* the same. But if $Av = \lambda v \iff P^{-1}BPv = \lambda v \iff B(Pv) = \lambda(Pv)$.

Lemma 174. Let $A \in M_{n\times n}(\mathbb{R})$. Assume $v_1, v_2, \ldots, v_n$ are eigenvectors of A. Then if $P = (v_1 \quad v_2 \cdots v_n)$ is invertible, then A is diagonalizable and $A = PDP^{-1}$ with D's leading diagonal entries being the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Proof. We can consider $P^{-1}AP = D \iff AP = PD$. Look at the k-th column vector of AP: $Av_k = \lambda_k v_k$. Look at the k-th column vector of PD: $P\lambda_k e_k$, which the diagonal identity matrix "extracts" out the eigenvectors one by one. (e_k is an identity matrix) □

3.5 Vector Space

3.5.1 Axioms and Examples

Definition 175. A ***vector space*** is a set V together with

- a binary operation: $+ : V \times V \to V$, $(v, \nu) \mapsto v + \nu$.
- a scalar multiplication: $\mathbb{R} \times V \to V$, $(\lambda, \nu) \mapsto \lambda\nu$.

such that:

(1) $(V, +)$ is an Abelian group

(2) $1 \cdot \nu = \nu$.

(3) $\forall \lambda, \mu \in \mathbb{R}, \nu \in V$: $\lambda(\mu\nu) = (\lambda\mu)\nu$

(4) $\forall \lambda \in \mathbb{R}, \nu, v \in V$: $\lambda(v + \nu) = \lambda v + \lambda\nu$

(5) $\forall \lambda, \mu \in \mathbb{R}, \nu \in V$, $(\lambda + \mu)\nu = \lambda\nu + \mu\nu$.

There are 9 axioms to be satisfied!

Example 176.

(1) Our "model": $\mathbb{R}^n$. Especially $\mathbb{R}$ is a vector space.

(2) $\mathbb{R}[x]_{\leq n} = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n | a_0, a_1, \ldots, a_n \in \mathbb{R}, x \in \mathbb{R}\}$.

(3) $M_{n,m}(\mathbb{R})$.

Lemma 177. V is vector space, $x \in V$, then

(1) $\forall n \in \mathbb{N}, nx = x + x + \cdots + x$.

(2) $0x = 0_V$.

(3) $(-1)x$ is additive inverse.

Proof.

(1) $1v = v \forall v \in V$. $2v = (1+1)v = 1v + 1v = v + v$. Induction!

(2) $0x = (0+0)x = 0x + 0x$. Since there is an inverse of $0x$ because addition is Abelian, add it on both sides:

$$0_V = 0x + (0x)^{-1} = 0x + 0x + (0x)^{-1} = 0x + 0_V = 0x.$$

(3) $0_V = 0x = (1 + (-1))x = 1x + (-1)x = x + (-1)x$, therefore $(-1)x = x^{-1}$ over addition.

$\square$

Definition 178. Let V be a vector space. A subset $U \subseteq V$ is called ***subspace*** if:

(1) If $x, y \in U$, $x + y \in U$. (Closure on addition)

(2) If $x \in U, \lambda \in \mathbb{R}$, then $\lambda x \in U$. (Closure on scalar multiplication)

(3) $0_V \in U$. (equivalent to saying $U \neq \varnothing$)

Note: $(U, +)$ is an Abelian group.

Example 179.

(1) Let $U \in \mathbb{R}^3, U = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} | x + y + z = 0\}$. It is geometrically a plane! (spanned by $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.) It "looks" like $\mathbb{R}^2$.

(2) Let $U \subseteq \mathbb{R}^{\mathbb{R}} := \{F | F : \mathbb{R} \mapsto \mathbb{R}\}$, $U := \{F | F(\pi) = 0\}$. (Be careful with the e, treat it as $0^f : x \mapsto 0$)

(3) $U := \{\lambda v | \lambda \in \mathbb{R}\} \subseteq V$.

(4) Every vector space has two *trivial* subspaces: itself and $\{0_V\}$.

If $U \subseteq V$ is a subspace which is neither itself nor $\{0_V\}$, then it is called a ***proper*** subspace.

Lemma 180. Let V be a vector space, $U, W \subseteq V$ are subspaces, then

(1) $U \cap W$ is a subspace.

(2) $U \cup W$ is *not* a subspace. (Unless $U \subseteq W$ and $W \subseteq U$.)

Proof. Exercise! (Write it out rigorously!) $\square$

Remark

- Concrete example for (2): let $U = \{\begin{pmatrix} x \\ 0 \end{pmatrix} | x \in \mathbb{R}\}, W = \{\begin{pmatrix} 0 \\ y \end{pmatrix} | y \in \mathbb{R}\}$. Both subspaces form a coordinate system, and their addition forms the coordinates on the coordinate system!

- $U + V := \{x + y | x \in U, y \in W\}$ is a subspace of V. (Exercise)

3.5.2 Spanning Sets

Definition 181. Let V be a vector space, $S \subseteq V$ be a subset. A ***linear combination*** of elements $v_1, v_2, \dots, v_n \in S$ is a vector $x \in V$:

$$x = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$$

where $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$.

Subsequently, all V denotes a vector space.

Lemma 182. Let $U \neq \varnothing, U \subseteq V$, then U is a subspace $\iff$ every linear combination of element of U is in U.

Proof. Exercise! $\square$

Definition 183. Let $S \subseteq V$ be a subset $S \neq \varnothing$,

$$\operatorname{Span} S := \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n | \lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}, v_1, v_2, \dots, v_n \in S\}.$$

In addition,

$$\operatorname{Span} \varnothing := \{0_V\}.$$

Proposition 184. Span S is a subspace of V.

Proof. Exercise! ☐

Definition 185. Set S is called the ***spanning set*** if Span $S = V$.

Example 186. Take $\text{Span}\{e_1, e_2, e_3\} = \mathbb{R}^3$, which forms the normal 3-D coordinate system we usually see. So $\{e_1, e_2, e_3\}$ is the spanning set of $\mathbb{R}^3$.

Lemma 187. $S \subseteq V$ subset, $S \subseteq U$ for some subspace $U \subseteq V$. Then span $S \subseteq U$.

Notation: Subsequently, $\subset$ and $\subseteq$ are used interchangeably, "strictly a subset of" is represented by $\subsetneq$, denoting "a proper subset", and depening on additional conditions, "a proper subspace".

Proof. U subspace $\Rightarrow$ closed under addition + scalar multiplication $\Rightarrow$ closed under taking linear combination. Then since $S \subseteq U \Rightarrow$ Span $S \subseteq U$ by definition. ☐

Warning: If S is a spanning set of V, it cannot be contained in any proper subspace of V. (If $S \subseteq U$ for some subspace $U \subset V$, then U is *not* closed under linear combinations of elements of S.)

Definition 188. A vector space is called ***finite-dimensional*** if it has a *finite spanning set*.

Notation: $\dim V < \infty$.

Note: If you want to prove that a vector space V is *not* finite-dimensional then it is not enough to find an infinite spanning set: you have to prove that no finite spanning set exists. Indeed, every vector space (except the zero vector space) has an infinite spanning set, e.g. the set $S = V$.

Definition 189. The ***dimension*** of a finite-dimensional vector space, written as $\dim V$, is the size of the smallest spanning set.

Example 190.

- Consider the vector space $\mathbb{R}[X]_{\leq d}$ of all polynomials of degree at most d. This is finite-dimensional, since the subset $S = \left\{1, X, X^2, \ldots, X^d\right\}$ is a spanning set.
- Consider the vector space $\mathbb{R}[X]$. It is not finite-dimensional. Suppose we pick a finite set $S = \{P_1, P_2, \ldots, P_n\} \subset \mathbb{R}[X]$. Each polynomial P_i has some degree d_i, and if we set $d = \max(d_1, d_2, \ldots, d_n)$, then S is contained in the proper subspace $\{\mathbb{R}[X]\}_{\leq d}$. So by the previous lemma, the subset S is not a spanning set.

3.5.3 Linear independence

Definition 191. A subset $L \subset V$ is called ***linearly dependent*** if we can find *distinct* vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in L$ and *non-zero* scalars $\lambda_1 \neq 0, \lambda_2 \neq 0, \ldots, \lambda_n \neq 0$ s.t.

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}_V.$$

If L is not linearly dependent we say that it is ***linearly independent***.

Note: If $L' \subset L$ and L' is linearly-dependent, then L is also linearly-dependent (just use the same $\mathbf{v}_i$'s and λ_i's). So if L is linearly-independent then any subset of L is also linearly-independent.

Example 192. If $\mathbf{0}_V \in L$ then L is linearly-dependent, because we can take any $\lambda \neq 0$ and observe that $\lambda \mathbf{0}_V = \mathbf{0}_V$.

Definition 193. A ***basis*** of a vector space is a linearly independent spanning set. The plural of basis is ***bases***.

Proposition 194. Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$ be a (finite) basis of vector space V. Then every vector in V can be written as a linear combination of elements in B, in a *unique* way. Conversely, any finite subset B with the above property is a basis.

Proof. Exercise! $\square$

Definition 195. If we find a (finite) basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then any vector $\mathbf{x} \in V$ can be uniquely expressed as

$$\mathbf{x} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n.$$

The scalars λ_i are called the ***coefficients of*** x ***with respect to*** B. Every vector $\mathbf{x} \in V$ corresponds to a unique set of coefficients $\lambda_1, \lambda_2, \ldots, \lambda_n$.

To put it more formally, the basis B allows us to define a function

$$F_B : \mathbb{R}^n \mapsto V$$
$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \mapsto \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n.$$

What Proposition 194 says is that the function F_B is a bijection. Note that F_B^{-1} sends a vector $\mathbf{x}$ to its coefficients with respect to B.

Lemma 196. Let $S \subset V$ be a spanning set, and suppose that S is not linearly independent. Then $\exists \mathbf{x} \in S$ s.t. $S' = S \setminus \{\mathbf{x}\}$ is still a spanning set.

Proof. Exercise! (Be careful that after removing an element from the spanning set, you need to show that it is not a spanning set of a subspace!)

Set $S' = S \setminus \{\mathbf{s}_1\}$. Then $\mathbf{s}_1 \in$ span S', and trivially $S' \subset$ span S', so $S \subset$ span S'. By Lemma 187, span $S \subset$ span $S' \Rightarrow V \subset$ span S'. This implies that span $S' = V$. □

Corollary 197. Any finite spanning set contains a basis.

Proof. Exercise! □

Corollary 198. Any finite-dimensional vector space V has a basis.

Proof. Exercise! □

Proposition 199. Let $S \subset V$ be a spanning set of V, and let $L = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a *finite linearly independent* subset of V, then $\exists T = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \subset S$, with the same size as L, s.t.

$$S' = (S \backslash T) \cup L$$

is a spanning set.

Proof. Exercise! □

Corollary 200. Let V be a finite-dimensional vector space, $S \subset V$ be a finite spanning set and $L \subset V$ be a linearly independent subset. Then L is finite and $|L| \leq |S|$.

Theorem 201. Let V be a finite-dimensional vector space with $\dim V = n$. Then any basis of V is finite and has size n.

Proof. Exercise! (Hint: use $\leq, \geq$ to deduce $=$.) □

Therefore, to check dimension, we just need to check if the spanning set is a basis and count its elements.

3.5.4 Dimension of Subspaces

Lemma 202. Assume $L \subseteq V$ linearly independent, $v \in V$, $v \notin \operatorname{span} L$, then $L \cup \{v\}$ still linearly independent.

Proof. Exercise! □

Lemma 203. If V is *not* finite dimensional ($\dim V = \infty$), then $\forall n \in \mathbb{N}, \exists L \subset V$ linearly independent subset s.t. $|L| = n$.

Proof. Exercise! (Hint: Induction!) Write it out formally! □

Lemma 204. Let $\dim V < \infty = n$. Then any linearly independent subset of size n is a basis.

Proof. Exercise! (Use contradiction!) $\square$

Lemma 205. Let $\dim V < \infty$, then any linearly independent subset is contained in a basis.

Proof. Exercise! $\square$

You can prove by adding linearly independent elements to the set, and this procedure is called ***basis extension***.

Example 206.

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^3$$

are linearly independent. Extend it to a basis.

$$\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \} = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} | x + 2z = 0 \} \subset \mathbb{R}^3.$$

So for instance, $v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \notin L$.

Proposition 207. Let $\dim V = n < \infty$, $U \subseteq V$ subspace. Then

(1) U is finite dimension

(2) $\dim U \leq \dim V$

(3) $\dim U = \dim V \Rightarrow U = V$.

Proof. Exercise! $\square$

Example 208. Possible subspaces of $\mathbb{R}^2$. (dim 0, 1, 2) ($\dim \mathbb{R}^2$) If $n = 0$: $\{0_V\}$ (0 vector space, *not* $\varnothing$!) If $n = 2 : \mathbb{R}^2$.

Proper subspaces have dimension $n = 1$ iff it has a one-dimensional *basis* $\{v\} = L$. $L = \{\lambda v | \lambda \in \mathbb{R}\}$, which is a line passing through the origin.

3.6 Linear Maps

Please visit subsection 3.3.2 for close link between the concepts introduced here and the properties related to the inverse of a matrix.

3.6.1 Definitions and Properties

Definition 209. Let U, V be vector spaces. A function $f : U \mapsto V$ is called ***linear*** if:

(1) $f(u+v) = f(u) + f(v)$

(2) $f(\lambda v) = \lambda f(v)$

where $u, v \in U$.

Example 210.

(1) Let $A \in M_{n,k}(\mathbb{R}), T_A : \mathbb{R}^k \to \mathbb{R}^n$, $x \mapsto Ax$.

(2) $0 : U \to V$, $u \mapsto 0_V$.

(3) $V = \mathbb{R}[x]$ (vector space of all polynomials). $D : \mathbb{R}[x] \to \mathbb{R}[x]$, $P \mapsto \frac{\mathrm{d}P}{\mathrm{d}x}(= P') =: D(P)$.

(4) $f : \mathbb{R}^2 \to \mathbb{R}$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^2 + y^2$.

Lemma 211. $f : U \to V$ is linear $\Rightarrow f(0_U) = 0_V$.

Proof. $f(0_U) = f(0_U + 0_U) = f(0_U) + f(0_U)$. Then by taking the inverse of $f(0_U)$ on both sides, i.e. minus $f(0_U)$ on both sides, because only addition is defined, so $f(0_U) = 0_V$. $\square$

Warning: The converse is not true!

Lemma 212. Let $f : U \to V, g : V \to W$ be linear, then $g \circ f : U \to W$ is also linear.

Proof. Exercise! $\square$

Definition 213. Let $f : U \to V$ be linear. Then the ***image*** of f is: Im $f := \{f(u) | u \in U\} \subseteq V$. The ***kernel*** of f is: Ker $f := \{u \in U | f(u) = 0_V\} \subseteq U$.

Lemma 214. Let $f : U \to V$ be linear. Then ker $f \subseteq U$ and Im $f \subseteq V$ are subspaces.

Proof. Exercise! □

Example 215. $A \in M_{1,3}(\mathbb{R}), A = (1, 1, 1)$. $T_A : \mathbb{R}^3 \to \mathbb{R}, x \mapsto Ax$ s.t.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x + y + z.$$

$$\ker T_A := \{v \in \mathbb{R}^3 | T_A v = 0_{\mathbb{R}}\} = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 | x + y + z = 0\}.$$

Lemma 216. A linear map $f : U \to V$ is injective $\iff$ ker $f = \{0_U\}$.

Proof. Exercise! □

Definition 217. The ***preimage*** of y is the set $f^{-1}(y) = \{x \in U | f(x) = y\}$.

Lemma 218. Let $f : U \to V$ linearly, $y \in V$ fixed. Suppose $x \in U$ s.t. $f(x) = y$. Then

$$f^{-1}(y) = \{x + v | v \in \ker f\}.$$

Proof. Exercise! (Warning: it is very hard to prove directly that two sets are equal! Instead, use $\subseteq$ and $\supseteq$ to derive $=$) Be careful with what you are proving, is it $=$ or $\subseteq$? For set, if proving $=$ directly, ensure that it is *bi-directional.* □

Example 219. $A \in M_{n,k}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^n$. $A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{R}^k \iff T_A\mathbf{x} = \mathbf{b}$. There are three possibilities:

(1) No solution: $\mathbf{b} \notin \operatorname{Im} T_A$.

(2) one solution: $\mathbf{b} \in \operatorname{Im} T_A$, $\ker T_A = \{0_{\mathbb{R}^k}\}$.

(3) ∞-many solution $\mathbf{b} \in \operatorname{Im} T_A$, $\dim \ker T_A \neq 0$.

Definition 220. Let $f : \mathbb{R}^k \to \mathbb{R}^n$ be linear. Then $f \equiv T_A$ for some matrix $A \in M_{n,k}(\mathbb{R})$.

Proof. Exercise! (It is constructive proof!) □

Proposition 221. Let $f : U \to V, g : U \to V$ be linear. $B = \{b_1, b_2, \dots, b_k\}$ basis of U. Assume $f(b_i) = g(b_i)\ \forall i$, then $f = g$.

Proposition 222. Let U, V vector spaces, $B = \{b_1, b_2, \dots, b_k\}$ basis of U, $\{v_1, v_2, \dots, v_k\} \subset V$. Then there exists a *unique linear* map s.t. $f(b_i) = v_i\ \forall i$.

Proof. Exercise! (There are three things to prove! $f(b_i) = v_i$, linear, and unique!) □

3.6.2 Isomorphism

Definition 223. A linear map $f : U \to V$ between two vector spaces is called an ***isomorphism*** if f is bijective. If there exists an isomorphism from U to V, we say that U is ***isomorphic*** to V, and write

$$U \cong V.$$

Note: f^{-1} is also an isomorphism, i.e. $U \cong V \iff V \cong U$. Since f is bijective, it is surjective $\Rightarrow$ $\operatorname{Im} f = V$, and injective $\Rightarrow$ $\ker f = \{\mathbf{0}_U\}$.

Example 224. $\mathbb{R}[X]_{\leq d} \cong \mathbb{R}^{d+1}, M_{2,2}(\mathbb{R}) \cong \mathbb{R}^4$.

Proposition 225. Let V be a vector space with $\dim V = n$. Then V is isomorphic to $\mathbb{R}^n$.

Proof. Exercise! (Remember to prove that the linear map is bijective!) □

Lemma 226. Let $f : U \to V$ be a linear map, $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$ be a basis for U. Let $C = \{f(\mathbf{b}_1), \dots, f(\mathbf{b}_k)\} \subset V$, then:

(i) C is a spanning set iff f is surjective

(ii) C is linearly independent iff f is injective

(iii) C is a basis iff f is an isomorphism.

Proof. Exercise! □

Corollary 227. If $U \cong V$ then $\dim U = \dim V$.

Proof. Exercise! □

Corollary 228. Let $f : U \to V$ be a linear map, $\dim U = \dim V$, then the following are equivalent:

(i) f is injective

(ii) f is surjective

(iii) f is an isomorphism.

Proof. Exercise! □

Corollary 229. If $f : \mathbb{R}^n \to V$ is an isomorphism, then the set

$$C = \{f(\mathbf{e}_1), f(\mathbf{e}_2), \dots, f(\mathbf{e}_n)\}$$

is a basis of V.

3.6.3 Rank-Nullity theorem

Definition 230. Let U and V be vector spaces, $f : U \to V$ be a linear map, then:

- the ***rank*** of f, written as rank f, is the dimension of Im f, i.e. $\dim \operatorname{Im} f$
- the ***nullity*** of f, written as null f, is the dimension of $\ker f$, i.e. $\dim \ker f$.

Theorem 231. Let U and V be vector spaces, $f : U \to V$ be a linear map, then

$$\text{rank } f + \text{ null } f = \dim U.$$

Proof. Exercise! □

Example 232.

(a)

$$\begin{aligned} D : \mathbb{R}[X]_{\leq n} &\to \mathbb{R}[X]_{\leq n-1} \\ P &\mapsto \frac{\mathrm{d}P}{\mathrm{d}X}. \end{aligned}$$

The kernel of D is the subspace

$$\ker D = \mathbb{R}[X]_{\leq 0} \subset \mathbb{R}[X]_{\leq n}$$

which has dimension 1. The image of D is the subspace

$$\operatorname{Im} D = \mathbb{R}[X]_{\leq n-1}$$

which has a dimension n. And indeed $\mathbb{R}[X]_{\leq n}$ has dimension $n+1$.

(b) Let $T_A : \mathbb{R}^3 \to \mathbb{R}^3$, $v \mapsto Av$, with $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Then

$$\ker T_A = \{v | Av = 0_{\mathbb{R}^3}\} = \text{ span } \{ \begin{pmatrix} 1 \\ -\frac{1}{2} \\ 0 \end{pmatrix} \}.$$

Hence the nullity of T_A is 1, and its rank is 2, since $\mathbb{R}^3$ has dimension 3.

(c) Let $A \in M_{n\times k}(\mathbb{R})$ be a matrix in RRE form, and $T_A : \mathbb{R}^k \to \mathbb{R}^n$ the associated linear map. Suppose that r of the columns of A containing a leading 1. Then from the definition of RRE form it's clear that Im T_A contains the first r standard basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_r \in \mathbb{R}^n$, and that these vectors span Im T_A, so the rank of T_A is r. Then the Rank-Nullity theorem says that

$$\text{null } T_A = \dim \{\mathbf{u} \in \mathbb{R}^k, A\mathbf{u} = \mathbf{0}\} = k - r.$$

This is the same as the number of 'free parameters' for this system of linear equations.

3.6.4 Linear Maps and Matrices

Suppose that U and V are arbitrary vector spaces of dimensions k and n, and that $f : U \to V$ is a linear map. We can associate a matrix with f by the following. Let

$$B = \{\mathbf{b}_1, \dots, \mathbf{b}_k\} \subset U \quad \text{and} \quad C = \{\mathbf{c}_1, \dots, \mathbf{c}_n\} \subset V$$

and be bases. We have seen that choosing bases gives us isomorphisms

$$F_B : \mathbb{R}^k \to U \quad \text{and} \quad F_C : \mathbb{R}^n \to V$$

by declaring that $F_B : \mathbf{e}_j \mapsto \mathbf{b}_j \forall j$, and $F_C : \mathbf{e}_i \mapsto \mathbf{c}_i \forall i$, and extending linearly. Now consider the map

$$F_C^{-1} \circ f \circ F_B : \mathbb{R}^k \to \mathbb{R}^n.$$

This map is linear, therefore it must be given by some matrix $A \in \text{Mat}_{n\times k}(\mathbb{R})$, i.e. there is an A s.t.

$$F_C^{-1} \circ f \circ F_B(v) = A\mathbf{v}$$

for all $\mathbf{v} \in \mathbb{R}^k$. This matrix A is called the ***matrix representing*** f ***with respect to*** B ***and*** C, and we write this as:

$${}_C[f]_B \quad \text{or} \quad [f]_B^C.$$

To compute this matrix ${}_C[f]_B$, recall that for any matrix the product $A\mathbf{e}_j$ is the jth column of A. So the jth column of the matrix ${}_C[f]_B$ is the vector

$$F_C^{-1} \circ f \circ F_B(\mathbf{e}_j) = F_C^{-1} \circ f(\mathbf{b}_j) \in \mathbb{R}^n.$$

The map $F_C^{-1} : V \to \mathbb{R}^n$ is the map that sends a vector $\mathbf{v}$ to the componenets of $\mathbf{v}$ with respect to C, so the procedure for finding ${}_C[f]_B$ is as follows:

- For each $j = 1, \dots, k$, take the jth basis vector $\mathbf{b}_j \in B$, and apply the map f to it to get a vector $f(\mathbf{b}_j) \in V$.
- Express each $f(\mathbf{b}_j)$ as a linear combination of the vectors in C

$$f(\mathbf{b}_j) = a_{1j}\mathbf{c}_1 + a_{2j}\mathbf{c}_2 + \cdots + a_{nj}\mathbf{c}_n$$

for some scalars $a_{1j}, \dots, a_{nj} \in \mathbb{R}$.

Then for each j, we have $F_C^{-1}(\mathbf{b}_j) = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$, so ${}_C[f]_B$ is the matrix (a_{ij}).

Example 233. Let $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map given by multiplying by the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

(Caution! Here, T_A is the "f" in definition! Don't confuse $\mathbb{R}^2 \to \mathbb{R}^2$ with the "$\mathbb{R}^k \to \mathbb{R}^n$" in the definition! On top of this, notice that T_A is a linear transformation, not the matrix itself! *One transformation can be represented using different matrices according to different bases!*)

If we pick the standard basis $B = C = \{\mathbf{e}_1, \mathbf{e}_2\}$ then

$$T_A(\mathbf{e}_1) = \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \mathbf{e}_1 + 3\mathbf{e}_2 \quad \text{and} \quad T_A(\mathbf{e}_2) = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2\mathbf{e}_1 + 4\mathbf{e}_2$$

so the matrix trpresenting T_A with respect to B and C is the matrix A itself. Clearly this is true, but we do not have to pick the standard bases! Let's take the same $A \in \mathrm{Mat}_{2\times 2}(\mathbb{R})$ as above, but choose bases

$$B' = \left\{ \mathbf{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^2$$

and

$$C' = \left\{ \mathbf{c}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{c}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^2.$$

Then

$$T_A(\mathbf{b}_1) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix} = 3\mathbf{c}_1 + \mathbf{c}_2$$

and

$$T_A(\mathbf{b}_2) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \mathbf{c}_1 + \mathbf{c}_2.$$

So the matrix representing T_A with respect to B' and C' is

$$_{C'}[T_A]_{B'} = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}.$$

This is indeed true since

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}.$$

It is important to understand that the linear map T_A has not changed here, all that's changed is the way in which we're choosing to write it down as a matrix.

Definition 234. Let V be a vector space, and let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be two bases for V. The ***change-of-basis matrix*** from B to C is the matrix

$$_C[\mathrm{id}_V]_B$$

that represents the identity map with respect to B and C. We usually denote the change-of-basis matrix by

$$_C P_B \quad \text{or} \quad P$$

if the choice of bases is clear.

Lemma 235. Let V be a vector space, let $B, C \subset V$ be two bases, and let $P = {}_C[\mathrm{id}_V]_B$ be the change-of-basis matrix. Pick any $\mathbf{x} \in V$. If the coefficients of $\mathbf{x}$ with respect to B are the vector $\mathbf{v} \in \mathbb{R}^n$, then the coefficients of $\mathbf{x}$ with respect to C are the vector $P\mathbf{v}$.

Proof. Exercise! $\square$

Example 236. Now consider the identity map $\mathrm{id} : \mathbb{R}^2 \to \mathbb{R}^2$. If $B, C \subset \mathbb{R}^2$ are both the standard basis then $_C[\mathrm{id}]_B$ is the identity matrix I_2. But if we take the bases B', C' from the example 233 then

$$\mathrm{id}(\mathbf{b}_1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \mathbf{c}_1 - \mathbf{c}_2 \quad \text{and} \quad \mathrm{id}(\mathbf{b}_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathbf{c}_1 - 2\mathbf{c}_2.$$

So the matrix representing id with respect to B' and C', i.e. change-of-basis matrix from B' to C', is

$$_{C'}[\mathrm{id}]_{B'} = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} = {}_{C'}P_{B'}.$$

Take the vector $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. If we write this using B' we get

$$\mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 0 \end{pmatrix} = -\mathbf{b}_1 + 2\mathbf{b}_2.$$

Thus the coefficients of $\mathbf{x}$ with respect to B are $F_B^{-1}(\mathbf{x}) = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$. The lemma claims that the coefficients of $\mathbf{x}$ with respect to C should be

$$P(F_B^{-1}(\mathbf{x})) = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix},$$

and indeed

$$\mathbf{c}_1 - 3\mathbf{c}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \mathbf{x}.$$

Proposition 237. Let $f : U \to V$ be a linear map. Let B and B' be two bases for U, and let $_BP_{B'}$ be the change-of-basis matrix between them. Similarly define $_{C'}P_C$ be the change-of-basis matrix between two bases for V, C' and C. Then the matrices representing f with respect to B and C or with respect to B' and C' are related by

$$_{C'}[f]_{B'} = {}_{C'}P_C \, {}_C[f]_B \, {}_BP_{B'}.$$

Proof. Exercise! □

Example 238. Let $D : \mathbb{R}[X]_{\leq 2} \to \mathbb{R}[X]_{\leq 1}$ be the differentiation map. If we choose bases $B = \{X^2, X, 1\}$ and $C = \{X, 1\}$ then the matrix representing D is

$$_C[D]_B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Now switch to the bases $B' = \{X^2+X+1, X+1, 1\}$ and $C' = \{X+1, 1\}$. To compute the change-of-basis matrix ${}_BP_{B'}$ we have to express each polynomial in B' as a linear combination of the elements of B — This is easy, and

$$ {}_BP_{B'} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}. $$

To compute the change-of-basis matrix ${}_{C'}P_C$ we have to express the polynomials in C in terms of the basis C'. This is also not hard, the matrix is

$$ {}_{C'}P_C = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}. $$

(This is the same problem as inverting the matrix ${}_CP_{C'}$.) The change-of-basis formula then says that

$$ \begin{aligned} {}_{C'}[D]_{B'} = {}_{C'}P_C\, {}_C[D]_B\, {}_BP_{B'} &= \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \end{pmatrix}. \end{aligned} $$

This is exactly the same matrix as the one if we were to obtain the matrix directly from B' and C'.

There is a special case of the change-of-basis formula which is important in its own right. Suppose that

$$ f : V \to V $$

is a linear map from V to itself. If we pick a single basis $B \subset V$ then we can write down f as the matrix ${}_B[f]_B$. Or we could choose another basis B' and write f down as the matrix ${}_{B'}[f]_{B'}$. If we let $P = {}_B[\mathrm{id}_V]_{B'}$ be the change-of-basis matrix from B' to B, then P^{-1} is the change-of-basis matrix from B to B', and the change-of-basis formula says that

$$ {}_{B'}[f]_{B'} = P^{-1} \left({}_B[f]_B\right) P. $$

For an apt application of this concept in action, check out subsection 2.3.2.

Remember that we say that two matrices M and N are ***similar*** if $M = P^{-1}NP$ for some invertible matrix P. So the matrices ${}_{B'}[f]_{B'}$ and ${}_{B}[f]_{B}$ are similar to each other. In addition, you can diagonalize a matrix A iff you can find a basis consisting of eigenvectors for A.

Observation on Similar Matrices

Notice that P needs to be *invertible* for the two matrices to be similar. Recall that P is invertible if $\neg\exists \mathbf{v} \neq \mathbf{0}$ s.t. $P\mathbf{v} = \mathbf{0}$. This is equivalent to saying, $\ker P = \{\mathbf{0}\}$. Thus nullity of P is 0. This means that P, as an $n \times n$ matrix, *cannot map different original bases to the same new basis!* In other words, if both bases are n dimensions, one of them cannot be *pseudo-n dimensions*—two vectors $\mathbf{v}_1, \mathbf{v}_2$ from one of the bases can be associated by $\mathbf{v}_1 = \lambda\mathbf{v}_2$, $\lambda \in \mathbb{R}$. In fact, one of the bases is therefore not even a basis, because $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly dependent.

For instance, although this might sound very ridiculous, but if a basis of $\mathbb{R}^2$, say A, is $\{(1,2),(2,4)\}$, this actually only has 1 dimension, and any attempt to doing change-of-basis matrix from other legitimate basis to this *pseudo-basis* will fail, because A is not even a basis, and it is one-dimensional.

In short, make sure that the chosen basis *is* a basis.

Chapter 4

Analysis

4.1 Sequence and Convergence

Definition 239. A ***sequence*** (of real numbers) is a map (function) $F : \mathbb{N} \mapsto \mathbb{R}$.

(Triangle Inequality) Theorem 240. Triangle inequality: $|x - y| \leq |x - z| + |z - y|$. (most common form: $|x + y| \leq |x| + |y|$)
"Reversed" triangle inequality: $||x| - |y|| \leq |x - y|$.

Proof. Assume $|x| \geq |y|$, and replace x with $x - y$ in $||x| - |y|| \leq |x - y|$:

$$|x| \leq |x - y| + |y| \Rightarrow ||x| - |y|| \leq |x| - |y| \leq |x - y|.$$

□

Definition 241. A sequence a_n is ***convergent*** if

$$\exists L \in \mathbb{R} \text{ s.t. } \forall \epsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N, |a_n - L| \leq \epsilon.$$

Definition 242. A sequence a_n is called ***divergent*** if

$$\forall L \in \mathbb{R}, \exists \epsilon > 0 \text{ s.t. } \forall N \in \mathbb{N}, \exists n \geq N \text{ s.t. } |a_n - L| \geq \epsilon.$$

Example 243.

(1) $a_n = \frac{1}{n} \to 0$. (Hint: Use Archimedian property: $\forall \epsilon > 0, \exists N \in \mathbb{N}$ s.t. $N > \frac{1}{\epsilon}$.)

(2) $a_n = C \to C$.

(3) $a_n = n$ is divergent.

(4) $a_n = (-1)^n$ is divergent. (Hint: applying triangle inequality, $|a_n - a_{n+1}| = |a_n - L + L - a_{n+1}| \leq |a_n - L| + |L - a_{n-1}|$.)

Definition 244. Let (a_n) be a sequence, $S \in \mathbb{N}$. The ***shift*** of (a_n) by S is the sequence $b_n := a_{n+S}$.

Lemma 245. Let (a_n) be a sequence, $S \in \mathbb{N}$, and $b_n := a_{n+S}$, then b_n converges $\iff$ a_n converges.

Proof. Exercise! □

Definition 246. A sequence is (strictly) decreasing if $\forall n \in \mathbb{N}, a_{n+1} \leq a_n$ (resp. $a_{n+1} < a_n$). A sequence is (strictly) increasing if $\forall n \in \mathbb{N}, a_{n+1} \geq a_n$ (resp. $a_{n+1} > a_n$).

Definition 247. Let (a_n) be a sequence. Then:

- (a_n) is ***bounded above*** if $\exists R_1 \in \mathbb{R}$ s.t. $a_n \leq R_1$.
- (a_n) is ***bounded below*** if $\exists R_2 \in \mathbb{R}$ s.t. $a_n \geq R_2$.
- (a_n) is ***bounded*** if $\exists R \in \mathbb{R}$ s.t. $|a_n| \leq R$. (Take $R = \max(|R_1|, |R_2|)$)

Definition 248. ***Supremum*** is the least upper bound, ***infimum*** is the biggest lower bound

(Completeness) Axiom 249. Let $S \subseteq \mathbb{R}, S \neq \varnothing$, bounded above (resp. below), then S has a supremum in $\mathbb{R}$ (resp. infimum).

Example 250. $S = \{s \in \mathbb{Q} : s^2 < 2\}$ does not have a supremum. Prove it!

Proposition 251. Let (a_n) be increasing and bounded above, then it is convergent, and $a_n \to \sup a_n$.

Proof. Exercise! □

Proposition 252. Let (a_n) be convergent, then a_n is bounded.

Proof. Exercise! □

Theorem 253. A bounded monotonic sequence is convergent.

Example 254. Consider $a_{n+1} = \sqrt{a_n + 6}, a_1 = 0$. Show that it is bounded (by 3) and monotonic. ($a_n = \sqrt{a_n a_n} < \sqrt{3a_n} = \sqrt{a_n + 2a_n} < \sqrt{a_n + 6} = a_{n+1}$.)

Proposition 255. Suppose $\exists L, M \in \mathbb{R}$ s.t. $a_n \to L$ and $a_n \to M$. Then $L = M$. (The limit is *unique*.)

Proof. Exercise! (using direct proof or contradiction) □

Theorem 256. Let (a_n) and (b_n) be sequences s.t. $a_n \to L, b_n \to M, \lambda \in \mathbb{R}$, then:

(1) $a_n + b_n \to L + M$

(2) $|a_n| \to |L|$

(3) $\lambda a_n \to \lambda L$

(4) $a_n b_n \to LM$

(5) If $M \neq 0, b_n \neq 0 \forall n, \frac{a_n}{b_n} \to \frac{L}{M}$.

Proof. Exercise! (the fourth one can be proven in two approaches) □

(Sandwich Test) Proposition 257. Assume $a_n \leq b_n \leq c_n \, \forall n, a_n \to L \wedge c_n \to L \Rightarrow b_n \to L$.

Proof. Exercise! □

Definition 258. Let (a_n) be a sequence, $(a_{F(n)})$ is a subsequence if $F : \mathbb{N} \mapsto \mathbb{N}$ is strictly increasing.

Proposition 259. Assume $a_n \to L$, then for any subsequence $a_{F(n)}$, $a_{F(n)} \to L$.

Proof. Exercise! □

Corollary 260. A sequence having (divergent) subsequences converging to different limits is divergent.

Proposition 261. Any sequence has a monotonic subsequence.

Proof. Exercise! (Recall *peak points*! Divide into two cases, where $S = \{n \,|\, a_m > a_n \, \forall m > n\}$.) □

Theorem 262. Any *bounded* sequence has convergent subsequence.

Proof. Exercise! □

Warninig: In general you cannot find the convergent subsequence explicitly.

Definition 263. Let (a_n) be a sequence. (a_n) is ***Cauchy*** if:

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \, \text{s.t.} \, \forall m, n > N, |a_m - a_n| < \epsilon.$$

Proposition 264. A Cauchy sequence is bounded.

Proof. Exercise! □

Proposition 265. Let (a_n) be a convergent sequence $\iff$ (a_n) is Cauchy.

Proof. Exercise! □

Warning: Cauchy $\iff$ convergent *in* $\mathbb{R}$ (and in $\mathbb{C}$), not in $\mathbb{Q}$!

4.2 Limits

Definition 266. Let $L \in \mathbb{R}$, $f : \mathbb{R} \mapsto \mathbb{R}$. $f(x) \to L$ as $x \to \infty$ iff

$$\forall \epsilon > 0, \exists R \in \mathbb{R} \text{ s.t. } x > R \Rightarrow |f(x) - L| < \epsilon.$$

Definition 267. Let $a \in \mathbb{R}$, $f : \mathbb{R} \mapsto \mathbb{R}$. $f(x) \to \infty$ as $x \to a$ iff

$$\forall M \in \mathbb{R}, \exists \delta > 0 \text{ s.t. } |x - a| < \delta \Rightarrow f(x) > M.$$

Definition 268. Let $L \in \mathbb{R}$, $f : \mathbb{R} \mapsto \mathbb{R}$. $f(x) \to \infty$ as $x \to \infty$ iff

$$\forall M \in R, \exists R \in \mathbb{R} \text{ s.t. } x > R \Rightarrow f(x) > M.$$

Definition 269. Let $a \in \mathbb{R}$, $F : (a, \infty) \mapsto \mathbb{R}$, $F \to L \in \mathbb{R}$ as $x \to a^+$ iff

$$\forall \epsilon > 0, \exists \delta > 0 \,\text{s.t.}\, x \in (a, a + \delta) \Rightarrow |F(x) - L| < \epsilon.$$

This is called the ***right-hand limit***. The pictorial illustration of this concept is as shown in Figure 4.1.

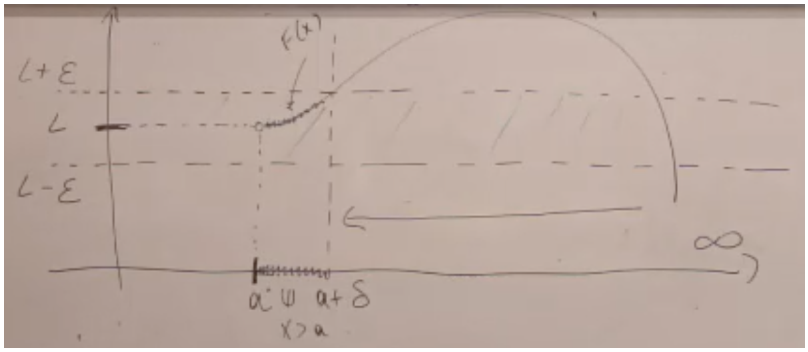


Figure 4.1: right limit illustration

Definition 270. Let $a \in \mathbb{R}$, $F : (-\infty, a) \mapsto \mathbb{R}$, $F \to L \in \mathbb{R}$ as $x \to a^-$ iff

$$\forall \epsilon > 0, \exists \delta > 0 \,\text{s.t.}\, x \in (a - \delta, a) \Rightarrow |F(x) - L| < \epsilon.$$

This is called the ***left-hand limit***. Graphical understanding is similar to that of right limit.

<u>Caution</u>: $F(a)$ does not need to be defined!

Lemma 271. Suppose $F : (a, b) \mapsto \mathbb{R}$, and assume that $\lim_{x\to a^+} = L_1$, $\lim_{x\to a^+} = L_2$, then $L_1 = L_2$.

Proof. Exercise! (Be careful to take $\delta := \min(\delta_1, \delta_2)$ instead of the maximum!) $\square$

(Limit) Definition 272. Let $a, L \in \mathbb{R}, f : \mathbb{R}\backslash\{a\} \mapsto \mathbb{R}$. $f(x) \to L$ as $x \to a$, or $\lim_{x\to a} f(x) = L$, iff

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - a| < \delta \Rightarrow |f(x) - L| < \epsilon.$$

Definition 273. Let $I \subset \mathbb{R}$ be an open iinterval and let $a \in I$ be a point. Take two functions $f, g : I \backslash \{a\} \mapsto \mathbb{R}$ s.t.

$$\lim_{x \to a} f(x) = L_1 \qquad \text{and} \qquad \lim_{x \to a} g(x) = L_2,$$

then

1. $\lim_{a \to 0}(f(x) + g(x)) = L_1 + L_2$
2. $\lim_{a \to 0}(f(x)g(x)) = L_1 L_2$
3. If $f(x) \neq 0 \ \forall x$, then $\lim_{a \to 0} \frac{1}{f(x)} = \frac{1}{L_1}$.

4.3 Continuity

Definition 274. Let $f : (b, c) \mapsto \mathbb{R}$, and pick $a \in (b, c)$. f is ***continuous at*** a iff

$$\lim_{x \to a} f(x) = f(a)$$

or the other way to say it is

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in \mathbb{R}[|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon].$$

Definition 275. Let $f : I \mapsto \mathbb{R}$, where $I \subset \mathbb{R}$ is either

- an interval (a, b) for some $a, b \in \mathbb{R}$, or
- an interval $(-\infty, b)$, or
- an interval (a, ∞), or
- $I = \mathbb{R}$,

we say that f is ***continuous everywhere***, or just ***continuous***, if f is continuous at $a \ \forall a \in I$. The four kinds of subsets $I \subset \mathbb{R}$ are called ***open intervals***.

Example 276. ***Rational functions*** $R(x) := \frac{P(x)}{Q(x)}$ are continuous, if P, Q polynomial, and $Q(X) \neq 0 \ \forall x$.

Strategy for $\epsilon - \delta$-proofs of continuity

1. Compute $f(a)$.
2. Look at $|f(x) - f(a)|$ to see how it can be controlled by $|x - a| < \delta$.
3. (This step may not be required.) Assume $\delta < C$ for some $C \in \mathbb{R}^+$ in order to control other possible factors of $|f(x) - f(a)|$ found in the previous step.
4. Find δ as a function of ϵ ($\delta(\epsilon)$) and write down the proof starting over from the begeinning.

Definition 277. f is ***discontinuous at*** a iff

$$\exists \epsilon > 0 \text{ s.t. } \forall \delta > 0, \exists x \in \mathbb{R}[|x - a| < \delta \wedge |f(x) - f(a)| \geq \epsilon].$$

While picking the particular x so that it satisfy the discontinuity condition, ensure that it satisfy for *all* δ.

Example 278. $F : \mathbb{R} \to \mathbb{R}$, $g(x) = 0$ when $x = 0$, and $g(x) = \sin \frac{1}{x}$ when $x \neq 0$. Prove that F is not continuous at 0.

Proposition 279. Let $I \subset \mathbb{R}$ and $J \subset \mathbb{R}$ be open intervals, $f : I \to F(I) \subset J \subset \mathbb{R}$, $g : J \to \mathbb{R}$ are continuous, then $g \circ f : I \to \mathbb{R}$ is continuous.

Proof. Since g is continuous at $b \in J$ so we can write $\forall \epsilon > 0, \exists \delta' > 0$ s.t. $\forall y \in J[|y - b| < \delta' \to |g(y) - g(b)| < \epsilon]$.

Similarly for f, since it is mapped from I to J we can write $\forall \delta' > 0, \exists \delta > 0$ s.t. $\forall x \in I[|x - a| < \delta \to |f(x) - f(a)| < \delta']$. We can write it as such because the codomain of f is the domain of g.

Combining the two statements, we can deduce that $\forall \epsilon > 0, \exists \delta' > 0, \delta > 0$ s.t. $\forall x \in I[|x - a| < \delta \to |f(x) - f(a)| < \delta' \to |g(f(x)) - g(f(a))| = |g \circ f(x) - g \circ f(a)| < \epsilon]$. □

4.3.1 Sequential Criterion for continuous functions

There is connection between (convergent) sequences and continuous function.

Proposition 280. Let $I \subset \mathbb{R}$, then $f : I \to \mathbb{R}$ is continuous $\iff$ $f(a_n) \to f(a)$ $\forall (a_n)$ s.t. $a_n \to a$.

Proof.

- "$\Rightarrow$": Since the function is continuous, we can deduce that $\forall \epsilon > 0, \exists \delta > 0$ s.t. $\forall x[|x - a| < \delta \to |a_n - a| < \epsilon]$.

  Since $a_n \to a$, we can write $\forall \epsilon' > 0\ \exists N \in \mathbb{N}$ s.t. $\forall n \geq N$, $|a_n - a| < \epsilon'$.

  Let $\epsilon' = \delta$, thus, $\forall \epsilon > 0, \exists \delta > 0, \exists N \in \mathbb{N}$ s.t. $\forall n \geq N, |a_n - a| < \delta \Rightarrow |f(a_n) - f(a)| < \epsilon$.

- "$\Leftarrow$": We can alternatively prove that, if $f(x)$ is not continuous at a, then $\exists a_n$ s.t. $a_n \to a$ s.t. $f(a_n)$ does not converge to $f(a)$.

  $f(x)$ is not continuous at a implies that $\exists \epsilon > 0$ s.t. $\forall \delta > 0$, $\exists x[|x - a| < \delta \wedge |f(x) - f(a)| \geq \epsilon]$.

  Let $\delta = \frac{1}{n}$, and substitute x with a_n, we obtain $\forall n > 0, |a_n - a| < \frac{1}{n}$. The sequence is thus convergent, and leave it as an exercise! While $f(a_n)$ does not converge to $f(a)$.

$\square$

Example 281.

$$f(x) = \begin{cases} x^2, & x \in \mathbb{Q} \\ -x^2, & x \notin \mathbb{Q} \end{cases} \quad \text{is continuous at } a \text{ iff } a = 0\ .$$

But $f(a_n) = a_n^2 \to a^2$, $f(b_n) = -b_n^2 \to -a^2$, and both are equal only if $a = 0$. Next, prove that it is continuous at $a = 0$.

Example 282.

$$f(x) = \begin{cases} \sin \dfrac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases} \quad \text{not continuous at } 0\ .$$

Challenge: Prove $\sin n$ diverges.

4.3.2 Continuous function on closed bounded interval

Definition 283. A function $f : K \to \mathbb{R}$, where $K = [b, c]$, is continuous if:

- f is continuous on (b, c).
- $\lim_{x \to c^-} f(x) = f(c)$.
- $\lim_{x \to b^+} f(x) = f(b)$.

Lemma 284. let $I \subset \mathbb{R}$ be open, $K \subset I$ be closed bounded. Then if $f : I \to \mathbb{R}$ continuous $\Rightarrow$ $f|_K : K \to \mathbb{R}$ continuous.

Definition 285. Let $S \subseteq \mathbb{R}$ be a subset, $f : S \to \mathbb{R}$ a function. f is called ***bounded*** $\iff$ $\exists R \in \mathbb{R}$ s.t. $|f(x)| \leq R \; \forall x \in S$.

Warning: Usually continuous functions on open interval are not bounded, e.g. $\frac{1}{x}$ on $(0, 1)$.

Proposition 286. Let $K = [b, c]$ closed bounded, $f : K \to \mathbb{R}$ continuous. Then f is bounded.

Proof. Assume f is not bounded $\iff$ $\forall n \in \mathbb{N}, \exists x_n \in K$ s.t. $|f(x_n)| > n$. Consider the sequence (x_n) on $K \Rightarrow (x_n)$ is bounded. By Bolzano Weierstraβ, $\exists$ convergent subsequence (x_{n_k}) s.t. $x_{n_k} \to L \in K = [b, c]$. Now by sequential criterion since f is continuous $f(x_{n_k}) \to f(L)$. But f is unbounded, therefore so is $f(x_{n_k})$, it cannot be convergent! $\square$

Definition 287. If $S \subseteq \mathbb{R}$, $f : S \to \mathbb{R}$. $\sup f := \sup \{f(x) | x \in \; S\}$, and $\inf f \; := \inf \{f(x) | x \in \; S\}$.

(Extreme Value) Theorem 288. Let K be a closed bounded interval, $f : K \to \mathbb{R}$ continuous. Then:

(1) $\exists x \in K$ s.t. $f(x) = \sup f$.

(2) $\exists y \in K$ s.t. $f(y) = \inf f$.

Proof. (1) Suppose that $f(x) \neq \sup f$, $\forall x \in K$. Define $g : K \to \mathbb{R}$, $g(x) := \frac{1}{\sup f - f(x)}$ well defined. So g is continuous on a bounded interval $\Rightarrow$ bounded. Since $\sup f$ is supremum, $\sup f - \frac{1}{n}$ is not an upper bound for any $n \in \mathbb{N}$. $\Rightarrow \forall n \in \mathbb{N}, \exists x_n \in K$ s.t. $f(x_n) > \sup f - \frac{1}{n} \iff \frac{1}{n} > \sup f - f(x_n) \iff n < \frac{1}{\sup f - f(x_n)} = g(x_n)$ is bounded!

Alternatively, let $a = \sup f$, then $\exists k \in K$ s.t. $a - \frac{1}{n} < f(k) \leq a$. This shows that $f(k) \to a$. Given that f is continuous, $\forall k_n \to k$, $f(k_n) \to f(k)$. Since limit is unique, $f(k) = a$.

(2) Largely identical process of proof.

□

(Intermediate Value) Theorem 289. Let $K = [b, c]$ be a *closed bounded* interval and let $f : K \to \mathbb{R}$ be *continuous*. Then:

1. If $f(b) \leq f(c)$, let $A \in \mathbb{R}$, $f(b) \leq A \leq f(c)$, then $\exists a \in [b, c]$ s.t. $f(a) = A$

2. If $f(b) \geq f(c)$, let $A \in \mathbb{R}$, $f(c) \leq A \leq f(b)$, then $\exists a \in [b, c]$ s.t. $f(a) = A$.

Proof. Exercise! (Hint: Use $\mathbb{R}$'s completeness axiom!) □

Technique: By utilizing the δ expression, one can always find a smaller/larger x that satisfies same inequality.

Corollary 290. If $f : [b, c] \to \mathbb{R}$ is continuous, and we have $f(b) < 0$ and $f(c) > 0$ (or vice versa), then $\exists a \in (b, c)$ s.t. $f(a) = 0$.

Corollary 291. Let $P : \mathbb{R} \to \mathbb{R}$ be any polynomial of odd degree. Then P has at least one root.

Proof. Define $P(x) = a_d x^d + Q(x)$, where $d \in \mathbb{N}$ is odd, $a_d \neq 0$, and $Q(x)$ is a polynomial of degree $\leq d - 1$.

Proceed from here! (Try to connect $Q(x)$ with $a_d x^d$ so that e.g. $P(x) > 0$ when $x > 0$, and vice versa) (Hint: sequential criterion!) □

Corollary 292. Let $K = [b, c]$ be a *closed bounded* interval, $f : K \to K$ be a continuous function. Then $\exists a \in K$ s.t. $f(a) = a$.

Proof. Exercise! (Think of a new function s.t. the previous corollary can be applied) □

Proposition 293. Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then $\exists c, d \in [a, b]$ s.t. the image $f([a, b])$ is the closed interval $[f(c), f(d)]$.

Proof. Exercise! (Hint: combine IVT and EVT!) □

Proposition 294. If $f : [a, b] \to \mathbb{R}$ or $f : \mathbb{R} \to \mathbb{R}$ is continuous, then f is injective $\iff$ it is strictly monotonic.

Proof. Exercise! □

Theorem 295. If $f : \mathbb{R} \to \mathbb{R}$ is continuous and injective, then $f^{-1} : f(\mathbb{R}) \to \mathbb{R}$ is continuous.

Proof. Exercise! □

4.3.3 Open, closed and compact sets

Definition 296. A set $S \subset \mathbb{R}$ is ***open*** iff

$$\forall x \in S, \exists \delta > 0 \text{ s.t. } (x - \delta, x + \delta) \subset S.$$

In other words, if x is a point of an open set S, then S has to contain every other point within a small *neighborhood* of x.

Proposition 297. Given a collection $\{S_\alpha\}$ of open subsets of $\mathbb{R}$, which may or may not be finite, the union $S = \bigcup_\alpha S_\alpha$ is open.

Proof. Exercise! □

Proposition 298. Given finitely many open sets $S_1, S_2, \ldots, S_n \subset \mathbb{R}$, the intersection $S = \bigcap_{i=1}^n S_i$ is open.

Proof. Exercise! □

Challenge: Prove that it is not true for infinitely many open sets.

Definition 299. A set $S \subset \mathbb{R}$ is ***closed*** iff

$$\forall \text{ sequence } (x_n) \subset S, x_n \to x \in \mathbb{R} \Longrightarrow x \in S.$$

In other words, the limit of any convergent subsequence of S must also be in S.

Definition 300. A set $S \subset \mathbb{R}$ is ***compact*** iff it is closed and bounded.

Example 301. Are the following sets closed?

- $\{\frac{1}{n} \,|\, n \in \mathbb{N}\}$,
- $[3, \infty)$,
- $\{40002\}$,
- $\mathbb{Q}$,
- $\mathbb{R}$.

Proposition 302. A set $S \subset \mathbb{R}$ is open iff its complement $T = \mathbb{R} \backslash S$ is closed.

Proof. Exercise! (The proof can show that, open set's complement must be closed, by definition, and not-closed set is not open.) □

The following two properties can be proved by utilizing the propositions 297 and 298.

Proposition 303. A union of finitely many closed sets is closed.

Proof. Exercise! □

Proposition 304. An intersection of arbitrarily many closed sets is closed.

Proof. Exercise! □

4.3.4 Uniform continuity and convergence

Definition 305. A function $f : S \to \mathbb{R}$ is said to be ***uniformly continuous*** iff

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x, y \in S[|x - y| < \delta \to |f(x) - f(y)| < \epsilon].$$

Definition 306. A function $f : S \to \mathbb{R}$ is not uniformly continuous iff

$$\exists \epsilon > 0 \text{ s.t. } \forall \delta > 0, \exists x, y \in S[|x - y| < \delta \wedge |f(x) - f(y)| \geq \epsilon].$$

Proposition 307. If $f : S \to \mathbb{R}$ is uniformly continuous, then it is continuous.

Proof. Exercise! □

Example 308. Determine if the following functions are uniformly continuous?

- $f(x) = ax + b$,
- $f(x) = x^2$,
- Define $f : (0, 1] \to \mathbb{R}$ by $f(x) = \frac{1}{x}$.

Proposition 309. If S is compact and $f : S \to \mathbb{R}$ is continuous, then f is uniformly continuous.

Proof. Suppose that f is not uniformly continuous. Then there must be an $\epsilon > 0$ s.t.

$$\forall \delta > 0, \exists x, y \in S \text{ s.t. } |x - y| < \delta \wedge |f(x) - f(y)| \geq \epsilon.$$

Now we look at $\forall \delta > 0, \exists x, y \in S$ s.t. $|x - y| < \delta$ to see if there is anything we can get. In general, this statement means that "you can always find two points in S s.t. they are arbitrarily close. (traditional sense of δ in statements about continuity is not applicable here, so should not confuse with the δ in the statement we extracted out.)

We take a sequence of points x_i, y_i with $|x_i - y_i| < \frac{1}{i}$ for all i. By Bolzano-Weierstrass, there is a subsequence (x_{i_j}) of the x_i which converges (since S is bounded) to some limit $x \in S$ (since S is closed). Then

$$\forall i, |x - y_{i_j}| \leq |x - x_{i_j}| + |x_{i_j} - y_{i_j}|$$

by the triangle inequality, and both terms on the right go to 0 as $j \to \infty$, so $y_{i_j} \to x$ as well. Since f is sequentially continuous at x, we know that

$$\lim_{j \to \infty} f(x_{i_j}) = f(x) = \lim_{j \to \infty} f(y_{i_j}).$$

So,

$$\exists N \in \mathbb{N} \text{ s.t. } \forall j \geq N \Big[|f(x_{i_j}) - f(x)| < \frac{\epsilon}{2} \wedge |f(x) - f(y_{i_j})| < \frac{\epsilon}{2}].$$

Combining these two triangle inequalities and we get

$$|f(x_{i_j}) - f(y_{i_j})| \leq |f(x_{i_j}) - f(x)| + |f(x) - f(y_{i_j})| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This contradicts with the assumption. $\square$

Definition 310. Let $f_1, f_2, \ldots : S \to \mathbb{R}$ be a sequence of functions defined on $S \subset \mathbb{R}$. We say that f_n converges ***pointwise*** to $f : S \to \mathbb{R}$ if

$$\forall x \in S, \forall \epsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } n \geq N \implies |f_n(x) - f(x)| < \epsilon.$$

We say that f_n converges ***uniformly*** to f if

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } \forall x \in S, n \geq N \implies |f_n(x) - f(x)| < \epsilon.$$

Example 311. Prove that $f_n(x) = x^n$ on $[0,1]$ converges pointwise to

$$f(x) = \begin{cases} 0, & 0 \leq x < 1 \\ 1, & x = 1 \end{cases},$$

but not uniformly.

Theorem 312. If a sequence of uniformly continuous functions $f_n : S \to \mathbb{R}$ converges uniformly to $f : S \to \mathbb{R}$, then f is uniformly continuous.

Proof. Exercise! □

Proposition 313. Let $S \subset \mathbb{R}$. If a sequence of continuous functions $f_n : S \to \mathbb{R}$ converges uniformly to $f : S \to \mathbb{R}$, then f is continuous.

Proof. Exercise! □

Definition 314. We say that a series

$$\sum_{i=1}^{\infty} f_i(x)$$

converges iff the sequence of partial sums

$$S_n(x) = \sum_{i=1}^{n} f_i(x)$$

converges, and it converges uniformly iff the sequence $S_n(x)$ converges uniformly.

(Weierstrass M-test) Theorem 315. Let $f_1, f_2, \ldots : S \to \mathbb{R}$ be a sequence of continuous functions, and suppose there are constants $M_1, M_2, \ldots$ s.t.

$$\forall i \in \mathbb{N}\ \forall x \in S, |f_i(x)| \leq M_i.$$

If $\sum_{i=1}^{\infty} M_i$ converges, then the series $\sum_{n=1}^{\infty} f_i(x)$ converges uniformly to a continuous function $g : S \to \mathbb{R}$.

Proof. Exercise! □

Example 316. Suppose for some $r > 0$ that the series $\sum_{i=0}^{\infty} a_i r^i$ converges absolutely, where the a_i are a sequence of real numbers. For all $i \geq 0$, we take

$$f_i(x) = a_i x^i, M_i = |a_i| r^i \Longrightarrow \forall x \in [-r, r], |f_i(x)| \leq M_i.$$

Since $\sum_i M_i$ converges, the Weierstrass M-test then tells us that the *power series*

$$\sum_{i=0}^{\infty} a_i x^i = \sum_{i=0}^{\infty} f_i(x)$$

converges uniformly to a continuous function on the interval $[-r, r]$.

Example 317. The series $f(x) = \sum_{i=0}^{\infty} \frac{\cos(13^i \pi x)}{2^i}$ converges uniformly on all of $\mathbb{R}$, since if we take $M_i = \frac{1}{2^i}$ for all i then

$$\left| \frac{\cos(13^i \pi x)}{2^i} \right| \leq M_i \text{ and } \sum_{i=0}^{\infty} M_i \text{ converges.}$$

This is one of a family of functions constructed by Weierstrass which are famously continuous on all of $\mathbb{R}$ but not differentiable anywhere.

4.4 Differentiability

Let $f : (b, c) \to \mathbb{R}, a \in (b, c)$. f is ***differentiable at*** a if

$$f'(a) := \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \text{ exists.}$$

$f'(a)$ is called the ***derivative*** at a.

Definition 318. Let $f : (b, c) \to \mathbb{R}, a \in (b, c), x - a =: h$. f is differentiable at a if

$$f'(a) := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \text{ exists.}$$

Proposition 319. If $f : I \to \mathbb{R}$ is differentiable, $I \subset \mathbb{R}$ is open, then f is continuous.

Proof. Exercise! □

Warning: The converse is not true! e.g. $f(x) = |x|$.

Proposition 320. Let $f : \mathbb{R} \to \mathbb{R}, I \subset \mathbb{R}$ be open, $a \in I$. Then the following statements are equivalent:

- f is differentiable at a.
- $\exists \lambda \in \mathbb{R}, \rho : I \to \mathbb{R}$ s.t. $\rho(a) = 0, \lim_{x \to a} \frac{\rho(x)}{x-a} = 0$, and

$$f(x) = f(a) + \lambda(x - a) + \rho(x).$$

Proof. Exercise! □

Proposition 321. Let $f, g : I \to \mathbb{R}$ be differentiable at $a \in I$, then:

1. $f + g$ differentiable and has derivative $f'(a) + g'(a)$
2. fg differentiable and has derivative $f'(a)g(a) + f(a)g'(a)$
3. If $f(x) \neq 0$ everywhere, then $\frac{1}{f}$ is differentiable with derivative $-\frac{f'(a)}{f^2(a)}$.

Proof. Exercise! □

Proposition 322. Let $f : I \to \mathbb{R}$ differentiable at a, $g : J \subset f(I) \to \mathbb{R}$ differentiable at $f(a)$, then $g \circ f$ is differentiable at a, and has derivative $g'(f(a))f'(a)$.

Proof. Exercise! □

Proposition 323. Let $f : I \to \mathbb{R}$ *strictly* increasing (or decreasing), differentiable at $a \in I$ s.t. $f'(a) \neq 0$. Then the inverse function $g : f(I) \to I$ exists and is differentiable with derivative

$$g'(b) = \frac{1}{f'(a)} = \frac{1}{f'(g(b))}.$$

Proof. Exercise! □

4.4.1 Extreme Values and Derivatives

Definition 324. Let $f : I \to \mathbb{R}$ be a function.

(1) f has a ***global maximum*** (resp. ***global minimum***) at $a \in I$, if $\forall x \in I, f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

(2) f has a ***local maximum*** (resp. ***local minimum***) at $a \in I$ if $\exists \epsilon > 0$ s.t. $\forall x \in I \cap (a - \epsilon, a + \epsilon)$, $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Warning: Such values do not necessarily exists!

Proposition 325. Let $f : (b, c) \to \mathbb{R}$ be differentiable at $a \in (b, c)$ and f has a local extremum. Then $f'(a) = 0$.

Proof. Exercise! □

Warning:

- Converse is not true! e.g. $f(x) = x^3$.
- Last proposition is not true for end points!

(Rolle) Theorem 326. Let $f : [a, b] \to \mathbb{R}$ be continuous on the closed bounded interval $[a, b]$ and differentiable on (a, b). Let $f(a) = f(b)$, then $\exists x \in (a, b)$ s.t. $f'(x) = 0$.

Proof. Exercise! □

(Mean Value Theorem) Theorem 327. Let $f : [a, b] \to \mathbb{R}$ be a continuous and differentiable on (a, b). Then $\exists x \in (a, b)$ s.t. $f'(x) = \frac{f(b)-f(a)}{b-a}$.

Proof. Exercise! (Hint: construct a new function and apply the Rolle's Theorem.) □

Example 328. Show that $f(x) = 4x^5 + x^3 + 7x - 2$ has exactly one root.

Corollary 329. Let $f : [a, b] \to \mathbb{R}$ be differentiable.

1. If $f'(x) \geq 0$ ($f'(x) > 0$), then f is (strictly) increasing.
2. If $f'(x) \leq 0$ ($f'(x) < 0$), then f is (strictly) decreasing.

Proof. Exercise! □

Corollary 330. Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable on (a, b). Then $f'(x) = 0 \ \forall x \in (a, b) \iff f$ is constant.

Technique: Think of using MVT in "the other way" round: arbitrarily choose two points, and the gradient of the line connecting the two points can be expressed using differentiation on one point in the interval.

Proposition 331. Let $f : I \to \mathbb{R}$ be differentiable. Let $L \in \mathbb{R}^+$ s.t. $|f'(x)| \leq L \ \forall x \in I$. Then $\forall x_1, x_2 \in I$,

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2|.$$

A function satisfying the above inequality is called ***Lipschitz continuous***.